THE AMERICAN MATHEMATICAL
MONTHLY

Volume 106, Number 2 February 1999

NOTICE TO AUTHORS

The MONTHLY publishes articles, as well as notes and other features, about mathematics and the profession. Its readers span a broad spectrum of mathematical interests, and include professional mathematicians as well as students of mathematics at all collegiate levels. Authors are invited to submit articles and notes that bring interesting mathematical ideas to a wide audience of MONTHLY readers.

The MONTHLY's readers expect a high standard of exposition; they expect articles to inform, stimulate, challenge, enlighten, and even entertain. MONTHLY articles are meant to be read, enjoyed, and discussed, rather than just archived. Articles may be expositions of old or new results, historical or biographical essays, speculations or definitive treatments, broad developments, or explorations of a single application. Novelty and generality are far less important than clarity of exposition and broad appeal. Appropriate figures, diagrams, and photographs are encouraged.

Notes are short, sharply focussed, and possibly informal. They are often gems that provide a new proof of an old theorem, a novel presentation of a familiar theme, or a lively discussion of a single issue.

Articles and Notes should be sent to the Editor:

ROGER A. HORN
1515 Mineral Square, Room 142
University of Utah
Salt Lake City, UT 84112

Please send your email address and 3 copies of the complete manuscript (including all figures with captions and lettering), typewritten on only one side of the paper. In addition, send one original copy of all figures without lettering, drawn carefully in black ink on separate sheets of paper.

Letters to the Editor on any topic are invited; please send to the MONTHLY's Utah office. Comments, criticisms, and suggestions for making the MONTHLY more lively, entertaining, and informative are welcome.

See the MONTHLY section of MAA Online for current information such as contents of issues, descriptive summaries of forthcoming articles, tips for authors, and preparation of manuscripts in TEX:

http://www.maa.org/

Proposed problems or solutions should be sent to:

DANIEL ULLMAN, MONTHLY Problems
Department of Mathematics
The George Washington University
2201 G Street, NW, Room 428A
Washington, DC 20052

Please send 2 copies of all problems/solutions material, typewritten on only one side of the paper.

Reprint permission:
DONALD ALBERS, Director of Publication

Advertising Correspondence:
Mr. JOE WATSON, Advertising Manager

Change of address, missing issues inquiries, and other subscription correspondence:
MAA Service Center
maahq@maa.org

All at the address:

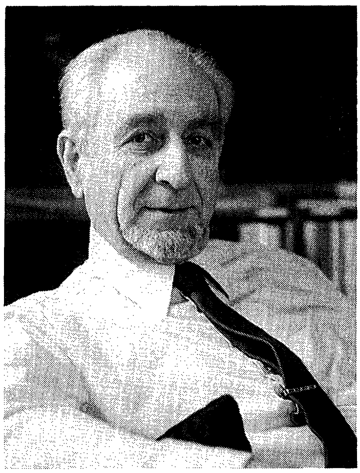
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036

Microfilm Editions: University Microfilms International, Serial Bid coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT. Second class postage paid at Washington, DC, and additional mailing offices. ***Postmaster***: Send address changes to the American Mathematical Monthly, Membership / Subscription Department, MAA, 1529 Eighteenth Street, N.W., Washington, DC, 20036-1385.

Yueh-Gin Gung and Dr. Charles Y. Hu Award for Distinguished Service to Leonard Gillman

Kenneth A. Ross



Many of us first became aware of Leonard Gillman in connection with "Gillman and Jerison," the classic 1960 monograph *Rings of Continuous Functions*. Len is a scholar and a perfectionist, and this shows clearly in this book. In particular, a lot of thought was put into creating the index, the most useful index I've ever seen. Chapter 16 is devoted to the Czech mathematician Katětov elegant characterization of the dimension of X in terms of $C^*(X)$. The authors completely reorganized and simplified the treatment. Katětov was astounded, as he had not believed the material could be simplified. I was honored when, at a meeting, Ed Hewitt introduced me to Gillman and Jerison in the same handshake. "Gillman and Jerison" was written while they were at Purdue. Later, Gillman became chairman at the University of Rochester where I had my first real job. Len is now Professor Emeritus at the University of Texas.

Actually, Len held a piano fellowship for five years at the Juilliard Graduate School before turning to mathematics. He has performed at five national meetings, three with Louis Rowen, cello, and two with William Browder, flute (AMS-MAA Presidents' Concert and Past-Presidents' Concert), and at several MAA Section meetings. All these concerts have been very well received.

Gillman's service to the profession goes back a long way. During the sixties, he became heavily involved at the national level: eight years as an MAA Visiting Lecturer and several years as a member of CUPM (Committee on the Undergraduate Program in Mathematics) and its subcommittees. CUPM was a big thing in those days, with its own national office (in Berkeley) and massive funding from NSF. He also spent two summers with SMSG (School Mathematics Study Group), writing "new math" materials, while at the same time criticizing the excesses of the new math thinking.

In 1973 Gillman became Treasurer of MAA. Like the Secretary, the Treasurer has a relatively long tenure. Much of what is accomplished makes the organization run smoothly, but does not make interesting reading. One such accomplishment was to have the Treasurer's Report no longer written in the Washington Office, but written by, well, the Treasurer, in order to replace accountant's jargon by readable English that Mr. and Ms. MAA member could understand. This was a nontrivial task.

A significant issue during this time concerned the MAA headquarters. The MAA had seriously outgrown the modest office space it had been renting. Al Willcox (Executive Director) found two connected townhouses on Eighteenth Street (numbers 1529 and 1527), for sale individually—or together along with the (unattached) carriage house, a package out of our league that would require a massive fund drive. Most members of the Executive and Finance Committees were ready to settle for just one house, but Gillman and a few others argued strongly for the package. Eventually this was the course adopted. The fund drive was successful, due in no small part to Gillman, who coaxed an endowment grant from James Vaughn and organized and directed a fund drive to name the conference room in memory of Ed Begle (Director of SMSG). The connected townhouses have since evolved into a small mathematics center.

Near the end of his term as Treasurer, Len became the leading advocate for conducting MAA's national elections by "approval voting." This scheme was adopted by the Board of Governors, though too late to affect Len, who was the last President elected under the old rules.

Here are two memorable events involving Len's Presidency. At the AMS Centennial Banquet at the January 1988 meeting in Atlanta, Len led the 1900 or so attendees in singing Happy Birthday Dear American Mathematical Society. Len's Retiring Presidential Address (Louisville, 1990), Teaching Programs That Work, discussed several remarkably successful programs, such as the one at SUNY Potsdam, Uri Treisman's work at Berkeley, and Jaime Escalante's at Garfield High School, focusing on their common features, and made a great hit.

Len has a knack for identifying good people and helping them move up into leadership positions. Among the two-year college people who caught his eye were Don Albers and Ann Watkins. Don is now MAA Associate Executive Director for Publications and Electronic Services. Ann served as co-editor with Bill Watkins of the *College Mathematics Journal*. Both Don and Ann have also served as MAA Vice Presidents.

Len's devotion to public service through his professional organizations apparently filtered down to his students. When he arrived at the University of Rochester, a young lady named Martha Jochnowitz arrived at the same time to begin graduate study in mathematics, and a math major named Doris Wood was starting her senior year. Both registered for his set-theory class. Doris ended the year as valedictorian of the graduating class, then went off to Yale for graduate study, and Len lost track of her. Martha married Chuck Siegel while completing her Ph.D. at Rochester and the rest is, as they say, history. Martha Siegel is now MAA Secretary. One day about a dozen years ago Len received a letter from Doris W. Schattschneider, editor of Mathematics Magazine, which included a handwritten note: "Do you still mark mathematical errors in red and errors in English in green?" Two editors later, the *Mathematics Magazine* editor was Martha Siegel! More recently, Doris was the Hedrick Lecturer at the 1995 mathfest in Burlington. Len had comparable influence on me and single-handedly brought me into MAA activities, which changed my life—for the better.

Len's Covection mathematical research involves a pleasant blend of topology, set theory, and analysis. Len's MAA publications include several journal articles, one of which won a Ford award; a booklet for high school students, *You'll Need Math*; and a manual for authors, *Writing Mathematics Well*.

University of Oregon, Eugene, OR 97403-1222
ross@math.uoregon.edu

Does Mathematics Need New Axioms?

Solomon Feferman

The question, "Does mathematics need new axioms?," is ambiguous in practically every respect.

- What do we mean by "mathematics"?
- What do we mean by "need"?
- What do we mean by "axioms"?

You might even ask, What do we mean by "does"?

Part of the ambiguity lies in the various points of view from which this question might be considered. The crudest difference lies between the point of view of the working mathematician and that of the logician concerned with the foundations of mathematics. Some logicians might protest this distinction since they identify themselves as (working) mathematicians who happen to specialize in mathematical logic. Certainly, modern logic has established itself as a very respectable branch of mathematics, and there are quite a few highly technical journals in logic, such as *The Journal of Symbolic Logic* and the *Annals of Pure and Applied Logic*, whose contents, from a cursory inspection, look just like those of other mathematical journals, setting subjects aside. Looking even closer, you can pick up a paper on, say, the semi-lattice of degrees of unsolvability or the model theory of fields and not see it as any different in general character from a paper on combinatorial graph theory or cohomology of groups; they belong to the same big frame of mind, so to speak. But if you pick up Gödel's paper on the incompleteness of axiom systems for mathematics, or his work and that of Cohen on the consistency and independence of the Axiom of Choice relative to the axioms of set theory, we're in a different frame of mind, because we are doing what Hilbert called *metamathematics*: the study of mathematics itself by the means of mathematical logic through its formalization in axiomatic systems. And it's that stance I want to distinguish from that of the mathematician working on analysis or algebra or topology or degrees of unsolvability, and so on. It's awkward to keep talking about the logician as metamathematician, and I won't keep qualifying it that way, but that's what I intend.

Though I won't at all neglect the viewpoint of the working mathematician, for most of this article I will be looking at the leading question from the point of view of the logician, and for a substantial part of that, from the perspective of one supremely important logician: Kurt Gödel. From the time of his stunning incompleteness results in 1931 to the end of his life, Gödel called for the pursuit of new axioms to settle undecided arithmetical problems. And from 1947 on, with the publication of his unusual article, "What is Cantor's continuum problem? **[12]**, he called in addition for the pursuit of new axioms to settle Cantor's famous conjecture about the cardinal number of the continuum. In both cases, he pointed primarily to schemes of higher infinity in set theory as the direction in which to seek these new principles. In recent years logicians have learned a great deal that is relevant to Gödel's program, but there is considerable disagreement about what conclusions to draw from their results. I'm far from unbiased in this respect, and

you'll see how I come out on these matters by the end of this essay, but I will try to give you a fair presentation of other positions along the way so you can decide for yourself which you favor.

The *Oxford English Dictionary* defines 'axiom' as used in Logic and Mathematics, by: "A self-evident proposition requiring no formal demonstration to prove its truth, but received and assented to as soon as mentioned." I think it's fair to say that something like this definition is the first thing we have in mind when we speak of axioms for mathematics: I'll call this the *ideal* sense of the word. It's surprising how far the meaning of axiom has been stretched from the ideal sense in practice, both by mathematicians and logicians. Some even take it to mean *an arbitrary assumption*, and so refuse to take seriously what purpose axioms are supposed to serve.

When the working mathematician speaks of axioms, he or she usually means those for some particular part of mathematics such as groups, rings, vector spaces, topological spaces, Hilbert spaces, etc. These axioms have nothing to do with self-evident propositions, nor are they arbitrary starting points. They are simply *definitions of kinds of structures* that have been recognized to recur in various mathematical situations. But they *act* as axioms in the sense that they provide a framework in which certain kinds of operations and lines of reasoning are appropriate whereas others are not. And once we run into a structure meeting one of these axiom systems—for example, a group associated with some equation or with a topological space—we can call on a vast body of previously established consequences for our further work. Without trying to argue this further, I take it that the value of these kinds of *structural axioms* for the organization of mathematical work is now indisputable. Moreover, we seem to keep coming up with new axioms of this sort, and I think the case can be made that they come up due to a continuing need to package and communicate our knowledge in digestible ways.

Now, in contrast to the working mathematician's structural axioms, when the logician speaks of axioms, he or she means, first of all, laws of valid reasoning that are supposed to apply to *all* parts of mathematics, and, secondly, axioms for such fundamental concepts as number, set, and function that underlie *all* mathematical concepts; I call the latter *foundational axioms*. I won't get into the question here of whether mathematics needs such axioms at all, and let the historical development of mathematics speak for that. Certainly, these correspond to such basic parts of our subject that they hardly need any mention at all in daily practice, and many mathematicians can carry on their entire work without calling on them even once. But that doesn't mean that they are not needed in the end to justify that practice, nor that they can safely be ignored in all situations. At any rate, I will take the necessity of foundational axioms for mathematics for granted in the following.

In particular, I will be concentrating on two axiom systems at conceptual extremes, the Dedekind-Peano Axioms for number theory and the Zermelo-Fraenkel axioms for set theory. I assume general familiarity with these, and so will skimp over the specifics of their formulations, which are in any case not important for the following; but I do have to say something about their development and the reasons for their acceptance. In both cases, one started with an informal "naive" system, which was later transformed into a formal system in the precise sense of metamathematics.

Dedekind's axioms [**3**] for the natural numbers $\mathbf{N} = \{0, 1, 2, \dots\}$ simply took the initial element 0 [Dedekind started with 1] and the successor operation $x \mapsto x'(= x + 1)$ as basic, with the evident axioms that 0 is not a successor and that successor is one-one. Induction was formulated set-theoretically, in the form: $\mathbf{N}$ is

the smallest set that contains 0 and is closed under the successor operation. This takes the informal notion of *arbitrary set* of natural numbers for granted, and in those terms the axioms are categorical and hence complete. Dedekind used the induction principle to show that one can define functions by simple recursion on **N**, i.e., where we prescribe how the function is defined at 0 and how it is defined at any successor element x' in terms of how it is defined at x; in particular this is used in the proof of categoricity. The functions of one or more arguments from **N** to **N** generated by explicit and simple recursive definition are nowadays called the *primitive recursive functions*.

Peano [**24**] made a first stab at adding axioms (to those of Dedekind) about which sets exist. He stated that every property determines a set, and then gave some closure conditions on properties. In the Peano axioms, induction is equivalent to the statement that any property of natural numbers that holds of 0 and is closed under successor holds of all natural numbers.

If one takes the notions of set or property as not needing any further analysis, then it seems to me that the Dedekind-Peano axioms come as close as anything we have to meeting the ideal dictionary sense of the word. But the metamathematical view is that *all* notions used in an axiomatic system and *all* assumptions concerning these must be fully spelled out. That is done by fixing a formal language for number theory, and taking for the properties in the induction principle just those expressed by a well-formed formula of that language. The resulting formal system nowadays is called Peano Arithmetic and denoted PA. In the formal language of PA, whose basic relation is that of equality, $=$, we need to add symbols for the operations $+$ and $\cdot$ to those for 0 and $'$, with their recursive defining equations as axioms; though $+$ and $\cdot$ are set-theoretically definable in terms of the latter by Dedekind's result, they are not definable from them in the first-order language used for formal arithmetic. However, Gödel showed in [**10**] that once we have 0, $'$, $+$, and $\cdot$, all primitive recursive functions are definable in PA.

Unlike the origin of the Dedekind-Peano axioms in a clear intuitive concept, Zermelo's axioms arose out of a need to give some sort of foundation to Cantor's revolutionary work in set theory, which many people regarded as problematic. In particular, Cantor made essential use of the Well-Ordering Principle (WO), according to which any set can be well-ordered, in order to establish various facts about cardinal arithmetic, in particular that for infinite cardinals κ, μ,

$$\kappa + \mu = \kappa \cdot \mu = \max(\kappa, \mu), \text{ while } \kappa < 2^{\kappa}.$$

Moreover, he used WO to show that the infinite cardinals can be laid out in a scale indexed by ordinals α,

$$\aleph_0 < \aleph_1 < \aleph_2 < \cdots < \aleph_\alpha < \aleph_{\alpha+1} < \cdots < \aleph_\lambda < \cdots,$$

where each $\aleph_{\alpha+1}$ is the least cardinal greater than $\aleph_\alpha$, and for limit λ, $\aleph_\lambda$ is the limit of all $\aleph_\alpha$ for $\alpha < \lambda$. This scale and the fact that $\aleph_0 < 2^{\aleph_0}$ immediately led to the conjecture known as the Continuum Hypothesis,

$$\text{(CH)} \qquad 2^{\aleph_0} = \aleph_1,$$

since $2^{\aleph_0}$ is the cardinal number of the continuum **R**. The extension of this conjecture to all α is called the Generalized Continuum Hypothesis,

$$\text{(GCH)} \qquad 2^{\aleph_\alpha} = \aleph_{\alpha+1}.$$

The question of justifying the Well-Ordering Principle was worrisome to Cantor. At first he argued that it is a "Law of Thought"; then he sought a proof of it on the basis of a more evident principle, but failed to come up with anything satisfactory. Such a principle was first offered in 1904 by Zermelo [**31**] in the form of the Axiom

of Choice (AC). Zermelo proved that AC implies WO; in fact, they are equivalent, but Zermelo argued that AC is evident in a way that WO is not. Following publication of this work, there were objections not only to the acceptance of AC but also to the correctness of his proof of the implication. In order to meet the latter objections, Zermelo introduced axioms in [**32**] that spelled out just which principles on sets were employed in his argument. These are the axioms of: Extensionality, Empty set, Unordered pair, Power set, Union, Infinity and Separation. The latter axiom says that for any *definite property* $P(x)$ of objects and any given set a, the set $b = \{x : x \in a \ \& \ P(x)\}$ also exists. This principle was objected to as being vague on what counts as a definite property, so, not long after, precise proposals were made independently by Weyl, Skolem, and Fraenkel to tie these down. Their proposals all essentially amount to taking for the definite properties just those expressed in the formal language for set theory, with basic symbols $=$ and $\in$. An additional modification was that Zermelo's axioms did not allow one to establish the existence of $\aleph_\alpha$ for infinite α; Fraenkel added his Replacement Axiom to accomplish this. As a formal system, the Zermelo-Fraenkel axioms are denoted ZF, and the same axioms to which AC is adjoined are denoted ZFC. A small point to note is that Zermelo allowed the existence of *urelements*, i. e., objects (other than the empty set) without elements. These have been dispensed with in ZF since they are not necessary for the foundations of set theory.

What was left unsettled by this development is an explanation of what, exactly, the Zermelo-Fraenkel axioms are axioms *for*. If they are to be considered to be axioms in the ideal, dictionary sense, they should be evident for some pre-axiomatic concept that we have in mind. The concept of *arbitrary set*, so to speak at large, which might first be offered as a candidate for this is unsatisfactory, because it seems to be an evident characteristic of this concept that for any property $P(x)$ the set of *all* x satisfying P exists. But as we know, this results in contradictions, the simplest being that due to Russell, using the property: $\neg(x \in x)$, where $\neg$ is the *negation* symbol. And it is just this sort of contradictory construction that Zermelo's Axiom of Separation avoids, by applying P only to elements of a "pre-existing" set a. What justifies that, but not the more general, contradictory concept of set? An answer was first offered by Zermelo [**33**] in 1930, in terms of what has since come to be called *the cumulative hierarchy of sets*. In this picture, sets are conceived of as being *built up* from below in stages, starting with the urelements at the lowest stage. Since we have dispensed with those, nowadays we simply start with the empty set, usually denoted 0. At each stage, we gather together all the sets obtained at preceding stages into a single set a. Then at the next stage we adjoin all members of the *power set* of a, $\wp(a)$, i.e., the *set of all subsets* of a. Finally, this process is iterated transfinitely. But to spell out what this model is, we need set theoretical notions themselves, as follows: the stages are indexed by ordinals. The set (or partial universe) of objects obtained at stage α is denoted V_α.

$$V_0 = 0,$$

$$V_{\alpha+1} = V_\alpha \cup \wp(V_\alpha), \quad \text{and}$$

$$V_\lambda = \text{the union of } \{V_\alpha : \alpha < \lambda\} \quad \text{for limit ordinals } \lambda.$$

It is argued by set-theorists nowadays that the axioms of ZFC are evident for the universe V of sets consisting of all objects in some V_α. But the intuition for that is a far cry from what leads one to accept the Dedekind-Peano axioms. Among other things, what this takes for granted is that there is an objective notion of arbitrary subset of a given set. This is the Platonistic conception of mathematics applied to

set theory, a conception which is philosophically controversial; we shall have more to say about that later on.

We return now to the origins of Gödel's program for new axioms in his 1931 paper [**10**] on the incompleteness for formal systems extending arithmetic. I want to remind you briefly of these results, for which we need some slightly technical notions. The simplest statements of number-theoretical interest are those of *purely universal form* $(\forall x) f(x) = 0$ and *purely existential form* $(\exists x) f(x) = 0$, where f is a primitive recursive function; these are dual forms under negation in the sense that $\neg(\forall x)f(x) = 0$ is equivalent to $(\exists x)g(x) = 0$ where $g(x)$ is 0 if $f(x) \neq 0$ and is 1 otherwise. A formal system S whose language contains that of PA is said to be *sound* for a class K of statements if whenever $S \vdash \phi$ (S proves ϕ) and $\phi \in K$, then ϕ is true in the natural numbers. It is easily shown that if S is *consistent* and contains PA (or even a weak fragment thereof) then it is sound for (purely) universal statements but (as Gödel showed) it need not be sound for existential statements. S is called 1-*consistent* just in case it is sound for existential statements. Note that 1-consistency implies consistency. (Gödel himself used a slightly stronger notion called *ω-consistency*.)

A system S is called *formally complete* if for each closed formula ϕ, either $S \vdash \phi$ or $S \vdash \neg \phi$. Hilbert had two fundamental conjectures about PA: that its consistency can be proved finitarily and that it is formally complete. Both conjectures were dashed by Kurt Gödel's incompleteness theorems of 1931 [**10**]. Moreover, they apply to (effectively presented) formal systems S extending PA much more generally. Gödel associated with each such system a purely universal statement θ_S, which expresses of itself, via its Gödel number, that it is not provable in S. Gödel's *first incompleteness theorem* has two parts. The first part tells us that if S is consistent then θ_S is indeed not provable in S, so by its very construction, it is true. The second part tells us that if S is 1-consistent then $\neg \theta_S$ is also not provable in S. For otherwise, being equivalent to an existential statement, if $\neg \theta_S$ were provable in S it would be true, contrary to the first part. Gödel's *second incompleteness theorem* tells us that the number-theoretic statement *Con*(S) expressing the consistency of S is not provable in S if S is consistent. This comes about by formalizing the proof of the first part of the first theorem. It follows that if S is a system in which all finitary reasoning can be formalized, then the general Hilbert finitary consistency program cannot be carried out for S. It is now generally accepted that all finitary reasoning can already be formalized in PA, if not in much weaker systems, and that's where Hilbert's finitary consistency program has its limits.

Not only were Gödel's results stunning, but also his own explanation of why they hold was surprising. This was given in a footnote that was apparently included in the paper [**10**] only as an afterthought, since it is numbered 48[a]. But it expressed a fundamental conviction of Gödel's which he reiterated throughout the rest of his life, and this conviction brings us close to the heart of our leading question. There is evidence that he thought such a view would be unacceptable to the Hilbert school, and that he must have hesitated to say anything of this sort at all. The footnote reads:

> ...the true reason for the incompleteness inherent in all formal systems of mathematics is that the formation of ever higher types can be continued into the transfinite...[since] the undecidable propositions constructed here become decidable whenever appropriate higher types are added. [**10**, p. 191]

However, the connection of incompleteness with set theory is not explained here;

the unstated reason is that the consistency of a system S can be proved in a system that uses variables for sets ranging over arbitrary subsets of the universe of discourse of S, by means of which a truth definition for the language of S can be introduced. Nothing more like footnote 48[a] was said by Gödel until the mid 1940s, by which time he was safely ensconced at the Institute for Advanced Study in Princeton, and Hilbert was dead and gone.

In the meantime, Gödel had established in **[11]** his second ground-breaking result, that if ZF is consistent then it remains consistent when we add AC and GCH. Gödel's method of proof for this was to produce a *new* cumulative hierarchy as a model of set theory, formally defined within set theory, by restricting the sets introduced at each stage to be all and only those subsets of the preceding stage which are definable in the language of set theory over that stage. The sets constructed in this way at stage α are denoted L_α, and their definition looks exactly like the sets V_α, except at the successor stages, where we take

$$L_{\alpha+1} = L_\alpha \cup Def(L_\alpha),$$

where $Def(a)$ for any set a is the *set of all definable subsets* b of a. A set is called *constructible* by Gödel if it belongs to some L_α; then L is used for the collection of all constructible sets. The so-called *Axiom of Constructibility* asserts that all sets are constructible, and is symbolized by $V = L$. This "axiom" served as a convenient intermediary in Gödel's relative consistency proof, as follows:

1. If ZF is consistent then ZF + $V = L$ is consistent.
2. ZF + $V = L \ \vdash$ AC & GCH.

Aside from the formal positioning of $V = L$ in 1 and 2, in what sense is its statement an acceptable axiom for set theory? At the time of his proof (circa 1938) Gödel stated that it provides a kind of natural completion of the axioms of set theory, since it ties down—in a way that ZF does not—exactly which sets we are talking about. But within a decade he was clearly rejecting it as an axiom, on the basis of a strongly Platonistic point of view of what set theory is supposed to be about. This position first emerged in an article on Russell's mathematical logic in 1944, but it was only stated forthrightly and with specific reference to open set-theoretical problems by Gödel in his 1947 article, "What is Cantor's continuum problem?" **[12]**, along the following lines:

1° Set theory is about a universe V of objects existing independently of human thoughts and constructions. It consists of the result of iterating into the transfinite the full power set operation, i.e., the operation of forming the set of arbitrary subsets of any given set. (So, on the basis of this, there is no reason to accept $V = L$, which says that all sets are introduced by successive definitions.)
2° Statements of set theory have a determinate truth value (in V). In particular, all axioms of ZFC are true in V.
3° So, also, CH has a determinate truth value. According to Gödel in **[12]** it is probably false.
4° Thus CH should be independent of ZFC. (Indeed, this was eventually demonstrated in 1963 by Paul Cohen **[2]**.)
5° And thus, in order to fix the position $\aleph_\alpha$ of $2^{\aleph_0}$ in the scale of alephs, we will [no doubt] need to add new axioms to ZFC.
6° These new axioms may be formulated and accepted by a direct extension of the informal reasoning that led us to accept ZFC in the first place. More

precisely:

> The simplest of these [new axioms]... assert the existence of [strongly] inaccessible numbers... $> \aleph_0$. [This] axiom, roughly speaking, means nothing else but that the totality of sets obtainable by exclusive use of the processes of formation of sets expressed in the other axioms forms again a set (and, therefore, a new basis for a further application of these processes). Other axioms of infinity have been formulated by P. Mahlo. ...these axioms show clearly, not only that the axiomatic system of set theory as known today is incomplete, but also that it can be supplemented without arbitrariness by new axioms which are only the natural continuation of those set up so far. [**12**, p. 520]

An uncountable cardinal is called (*strongly*) *inaccessible* if it is closed under exponentiation and limits of smaller cardinals. It follows that if κ is inaccessible then V_κ is a model of the ZFC axioms. Thus, if it is assumed that there exists an inaccessible cardinal then the consistency of ZFC, Con(ZFC), is a consequence and so, by Gödel's second incompleteness theorem, it is not provable in ZFC (if ZFC is consistent). Similarly, if one assumes there are a certain number of inaccessible cardinals, then one will not be able to prove the existence of larger inaccessibles. The Mahlo axioms assert the existence, to begin with, of arbitrarily large inaccessibles, and then of arbitrarily large inaccessible fixed points of the enumeration of the inaccessibles, and so on, iterated into the transfinite. An informal way of justifying their existence, and, indeed, of infinite cardinals at all, is by reference to "Cantor's Absolute": the universe of all sets is beyond being captured by any closure condition on sets; instead, any such condition always closes off at a set. A bit more explicitly, whatever closure property P one recognizes to be satisfied by the universe V of all sets, there will exist arbitrarily large κ for which V_κ satisfies P. Formal versions of this, introduced by Azriel Levy [**20**] and Paul Bernays [**1**], are called *Reflection Principles* in set theory. They are behind Gödel's reason for saying that we are led to new axioms, such as those of Mahlo type, "without arbitrariness" and as a "natural continuation" of those axioms previously accepted. But, he continued [from the preceding quote],

> [a]s for the continuum problem, there is little hope of solving it by means of those axioms of infinity which can be set up on the basis of principles known today... [ibid.]

The reason is that the Mahlo axioms are consistent with $V = L$, and since GCH is true in L, and Gödel believed CH to be false, its falsity could not be proved in this way. "In the face of this," he continued on,

> ... probably there exist other [axioms] based on hitherto unknown principles...which a more profound understanding of the concepts underlying logic and mathematics would enable us to recognize as implied by these concepts. [ibid.]

A candidate for a larger kind of cardinal had in fact been suggested some time before, by Stanislaw Ulam, in 1930. Ulam called an uncountable cardinal κ *measurable* if there exists a two-valued κ-additive measure on $\wp(\kappa)$. Not much was known about the strength of this until 1961, when Dana Scott proved in [**25**] that the existence (MC) of measurable cardinals implies $V \neq L$, so MC then became a

viable possibility to settle CH. A few years later, Alfred Tarski with his students William Hanf and H. Jerome Keisler proved ([**16**], [**19**]) that if κ is a measurable cardinal then it is very large, since V_κ satisfies the axioms of Mahlo type and other powerful axioms of infinity. Their work led further to a notion of *strongly compact cardinal*, whose existence was shown to imply the existence of measurable cardinals. But then both Tarski and Gödel had qualms about the assumption of the existence of such enormous cardinals. To quote Tarski:

> the belief in the existence of inaccessible cardinals...(and even of arbitrarily large cardinals of this kind) seems to be a natural consequence of basic intuitions underlying the "naive" set theory and referring to what can be called "Cantor's absolute." On the contrary, we see at this moment no cogent intuitive reasons which could induce us to believe in the existence of [strongly compact] cardinals, or which at least would make it very plausible that the hypothesis stating the existence of such cardinals is consistent with familiar axiom systems of set theory. [**27**, p. 134]

In his 1964 revision [**13**] of his 1947 article, Gödel seconded this view of Tarski's in full, but then added:

> However, [the new axioms] are supported by rather strong argument from *analogy*... ([**13**, p. 264, ftn. 20], italics mine)

Moreover, Gödel had already proposed in 1947 still another kind of argument that might lead one to accept certain statements as new axioms, even though they do not rest on the same kind of evidence that led one to accept ZFC in the first place, to wit:

> [Finally, we may look for axioms which are] so abundant in their verifiable consequences...that *quite irrespective of their intrinsic necessity* they would have to be assumed in the same sense as any well-established physical theory. ([**12**, p. 521], italics mine)

Higher axioms of infinity, or so-called "large cardinals" in set theory have been the subject of intensive investigation since the 1960s and many new kinds of cardinals with special set-theoretical properties have emerged in these studies.[1] A complicated web of relationships has been established, as witnessed by charts to be found in the recent book by Aki Kanamori, *The Higher Infinite* [**17**, p. 471], and the earlier expository article by Kanamori and Menachem Magidor [**18**]. A rough distinction is made between "small" large cardinals, and "large" large cardinals, according to whether they are weaker or stronger, in some logical measure or other, than measurable cardinals. Attempts to justify acceptance of both kinds of cardinals have been made by set theorists involved in this development. The philosopher, Penelope Maddy, in two interesting articles called "Believing the axioms," analyzed the various kinds of arguments for these and other kinds of strong axioms and summarized the evidence for them [**21**]. Broadly speaking, the

[1]The elaboration of this subject has almost outrun the names that have been introduced for various large cardinal notions, witness (in roughly increasing order of strength): "inaccessible," "Mahlo," "weakly compact," "indescribable," "subtle," "ineffable," "Ramsey," "measurable," "strong," "Woodin," "superstrong," "strongly compact," "supercompact," "almost huge," "huge," and "super-huge."

arguments are classified as being based on intrinsic or extrinsic reasons. The above-mentioned reflection principles are examples of intrinsic reasons, but these do not take us beyond the "small" large cardinals. Among the extrinsic reasons for going higher are that the assumption of "large" large cardinals has been fruitful—through the dazzling work of Solovay, Martin, Foreman, Magidor, Shelah, Steel, Woodin and others—in extending "standard" properties of Borel and analytic subsets of the continuum, such as Lebesgue measurability, the Baire property, the perfect subset property, determinateness of associated infinite games, etc. to substantially larger classes.

But the striking thing, despite all this progress, is that contrary to Gödel's hopes, the Continuum Hypothesis is *still* undecided by these further axioms, since it has been shown to be independent of all remotely plausible axioms of infinity, including MC, that have been considered so far (assuming their consistency)[2]. That may lead one to raise doubts not only about Gödel's program but also about its very presumptions. Is CH a definite problem as Gödel and many current set-theorists believe? Is the continuum itself a definite mathematical entity? If it has only *Platonic* existence, how can we access its properties? Alternatively, one might argue that the continuum has *physical* existence in space and/or time. But then one must ask whether the *mathematical structure* of the real number system can be identified with the *physical structure*, or whether it is instead simply an *idealized mathematical model* of the latter, much as the laws of physics formulated in mathematical terms are highly idealized models of aspects of physical reality. (Hermann Weyl raised just such questions in his 1918 monograph *Das Kontinuum*, **[29]**.) But even if we grant some kind of independent existence, abstract or physical, to the continuum, in order to formulate CH we need to refer to arbitrary *subsets of the continuum* and possible mappings between them, and then we are dealing with objects of a higher level of abstraction, the nature of whose existence is even more problematic than that of the continuum. Here we are skirting deep philosophical waters; let us retreat from them for the moment.

While Gödel's program to find new axioms to settle CH has not been realized, what about the origins of his program in the incompleteness results for number theory? As we saw, throughout his life Gödel said we would need new, ever-stronger set-theoretical axioms to settle open arithmetical problems of even the simplest, purely universal, form—problems he called of *Goldbach type*. Indeed, the Goldbach conjecture can be written in that form. But the incompleteness theorem by itself gives no evidence that any open arithmetical problems—or, equivalently, finite combinatorial problems—of *mathematical interest* will require new such axioms. I emphasize the 'mathematical interest', because Gödel's own examples of undecidable statements for each consistent S extending PA were of two kinds: the first, θ_S, cooked up by a diagonal construction in order to establish incompleteness and evidently true by the very theorem that it is used to prove, and the second, $\mathrm{Con}(S)$, of definite *metamathematical interest*, but not of mathematical interest in the ordinary sense of the word. Beginning in the mid-1970s, logicians began trying to rectify this situation by producing finite combinatorial statements of prima-facie mathematical interest that are independent of such S. The first example was provided by Jeff Paris and Leo Harrington who showed in **[23]** that a modified form (PH) of the finite Ramsey theorem concerning existence of homogeneous sets for certain kinds of partitions is not provable in PA. PH is recognized to be true as

[2]Cf. Martin **[22]**; the situation reported there in 1976 remains unchanged to date.

a simple consequence of the infinite Ramsey theorem; its independence rests on showing that PH implies Con(PA); in fact PH is equivalent to 1-Con(PA). Moving up to a stronger system, a few years later, Harvey Friedman, Ken McAloon, and Stephen Simpson produced a finite combinatorial version FGP of the Galvin-Prikry theorem GP that is independent of the Feferman-Schütte system of predicative analysis (call it FS for present purposes)[3]. It happens that GP is itself a considerable strengthening of the infinite Ramsey Theorem, and FGP has certain analogies to PH. Again, this finitary version FGP is proved to be true as a simple consequence of GP, while its independence rests on showing that it implies Con(FS); in fact, FGP is equivalent to the 1-consistency of predicative analysis. Further results of this type have been obtained by these researchers and others for still stronger systems of analysis[4]. While in each case, the statement ϕ shown independent of S is equivalent to its 1-consistency, the argument for the truth of ϕ is by ordinary mathematical reasoning.

For some years, Friedman has been trying to go much farther, by producing mathematically perspicuous finite combinatorial statements ϕ whose proof requires the existence of many Mahlo axioms or even stronger axioms of infinity and has come up with various candidates for that ([7] contains the latest work in this direction). From the point of view of metamathematics, this kind of result is of the same character as the earlier work just mentioned; that is, for certain very strong systems S of set theory, the ϕ produced is equivalent to the 1-consistency of S. But the conclusion to be drawn is not nearly as clear as for the earlier work, since the truth of ϕ is now *not* a result of ordinary mathematical reasoning, but depends essentially on acceptance of 1-Con(S). It is begging the question to claim this shows we need axioms of large cardinals in order to settle the truth of such ϕ, since our *only* reason for accepting that truth lies in our belief in the 1-consistency of those axioms. However plausible we might find that, perhaps by some sort of picture we can form of the models of such axioms, it doesn't follow that we should accept *those axioms themselves* as first-class mathematical principles. Finally, we must take note of the fact that up to now, *no previously formulated open problem* from number theory or finite combinatorics, such as the Goldbach conjecture or the Riemann Hypothesis or the twin prime conjecture or the P = NP problem, is known to be independent of the kinds of formal systems we have been talking about, not even of PA. If such were established in the same way as the examples (PH, FGP, etc.) mentioned above, then their truth would at the same time be verified. I think it is more likely, as has been demonstrated in the case of the Fermat "last theorem," that the truth of these will eventually be settled—if at all—by ordinary mathematical reasoning without any passage through metamathematics, and that only afterward might we see just which basic axiomatic principles are required for their proofs.

[3]Friedman, McAloon, and Simpson work with a system ATR_0 which is shown to be proof-theoretically equivalent to the FS system of ramified analysis up to the ordinal Γ_0. Friedman later found a finite version of Kruskal's theorem KT which is independent of ATR_0. The infinitary theorem KT, a staple of graph-theoretic combinatorics, asserts the well-quasi-ordering of the embeddability relation between finite trees. Friedman's work in this respect is reported in [**26**].

[4]The systems involved and associated independent statements are more complicated to explain and would go beyond the scope of this article to do so, but at least one result is worth indicating in connection with footnote 3. Friedman found an extended version EKT of KT which is independent of the impredicative Π_1^1 comprehension principle in analysis (cf. [**26**]). EKT later turned out to have close mathematical and metamathematical relationships with the graph minor theorem of Robertson and Seymour, as shown in [9].

Moving beyond the domains of arithmetic and finite combinatorics, what is the evidence that we might need new axioms for everyday mathematics? Here it is certainly the case that various parts of descriptive set theory have been shown to require higher axioms of infinity, in some cases well beyond the range of "small" large cardinals. But again we are in a question-begging situation, since our belief in the truth of these new results depends essentially on our belief in the consistency or correctness to some extent or other of these "higher" statements. Also, I think it is fair to say that these kinds of results are at the margin of ordinary mathematics, that is of what mathematicians deal with in daily practice.[5] What is *not* at the margin can be readily formalized within ZFC, and in fact in much weaker systems, as has been demonstrated by many case studies in recent years.

Let's look more specifically at the part of mathematics that is indispensable to scientific applications, which clearly includes vast tracts of analysis, among other subjects. One of the arguments for accepting any set theory at all, if one is not a Platonist, has been advanced by the philosophers Willard van Orman Quine and Hilary Putnam, along the lines that some set theory is necessary for the foundations of analysis, and that the resulting mathematics is justified by its essential and successful use in established physical theory. But this argument is undermined by a series of case studies, beginning with that of Hermann Weyl in 1918, in his famous monograph *Das Kontinuum* **[29]**, in which he showed in principle how all of nineteenth-century analysis of piecewise continuous functions could be formalized in a system S reducible to PA; this has been continued since the mid-70s with work by Gaisi Takeuti, Harvey Friedman, Stephen Simpson, and myself among others, to extend this to substantial portions of twentieth-century analysis including much of measure theory and functional analysis. As a result of these studies, I have come to conjecture that practically all scientifically applicable mathematics can be formalized in systems reducible to PA, or, as I have sloganized it in **[4]**: *a little bit goes a long way*. Against this, I have learned of a couple of cases in some approaches to the foundations of quantum field theory where it appears one must go beyond the resources of PA; but the physical theories that require such additional strength are rather speculative. In any case, the mathematics needed for these cases can be carried out in relatively weak subsystems of impredicative analysis, even if PA does not suffice. I am not by any means arguing that everyday mathematical practice should be restricted to working in such subsystems. The instrumental value of "higher" and less restricted set-theoretical concepts and principles is undeniable. The main concern here is, rather, to see: *what*, *fundamentally*, *is needed for what*?

To conclude, I hope I have given you some food for thought that will help you come to your own conclusions about whether questions like the Continuum Hypothesis are determinate, and, if so, what is going to settle them, given that present axioms are insufficient. At the beginning of this piece I promised to tell you my own views of these matters. By now, you have probably guessed what these are, but let me say them out loud: I am convinced that the Continuum Hypothesis is an inherently vague problem that *no* new axiom will settle in a convincingly definite way[6]. Moreover, I think the Platonistic philosophy of mathematics that is

[5]For an opposite point of view and beautiful exposition of the need for new axioms in that respect, cf. Woodin [30].

[6]CH is just the most prominent example of many set-theoretical statements that I consider to be inherently vague. Of course, one may reason confidently *within* set theory (e.g., in ZFC) about such statements *as if* they had a definite meaning.

currently claimed to justify set theory and mathematics more generally is thoroughly unsatisfactory and that some other philosophy grounded in inter-subjective *human* conceptions will have to be sought to explain the apparent objectivity of mathematics. Finally, I see no evidence for the practical need for new axioms to settle open arithmetical and finite combinatorial problems. The example of the solution of the Fermat problem shows that we just have to work harder with the basic axioms at hand. However, there is considerable theoretical interest for logicians to try to explain what new axioms *ought to be accepted* if one has already accepted a given body of principles, much as Gödel thought the axioms of inaccessibles and Mahlo cardinals *ought to be accepted* once one has accepted the Zermelo-Fraenkel axioms. In fact this is something I've worked on in different ways for over thirty years; during the last year I have arrived at what I think is the most satisfactory general formulation of that idea, in what I call the *unfolding of a schematic formal system* **[5]**. And this returns in an essential respect to the original "naive" schematic formulation of principles such as induction in number theory and separation in set theory, in their use of the pre-theoretic notion of arbitrary "definite" property. That is in closer accord with everyday practice, where such principles are taken in an *open-ended* way, without advance restriction on what specific language they are formulated in. But we can systematically enlarge what we regard as meaningful in a given subject, by using those very principles in a kind of feed-back way, for example in the use of induction to prove that a function or predicate of natural numbers defined implicitly by recursion equations is total and thus can be added to our language. There are already some definitive results for specific systems on what can be obtained by the unfolding process, in joint work with Thomas Strahm **[6]**, with a host of new and interesting problems waiting to be tackled. But that's another story for another occasion.

ACKNOWLEDGMENTS. Text of an invited AMS-MAA lecture, Joint Annual Meeting, San Diego, Jan. 10, 1997.

REFERENCES

1. P. Bernays, Zur Frage der Unendlichkeitsschemata in der axiomatischen Mengenlehre, in (Y. Bar-Hillel, et al., eds.) *Essays on the Foundations of Mathematics*, Magnes Press, Jerusalem, 1961, pp. 3–49. (English translation in (G. H. Müller, ed.) *Sets and Classes*, North-Holland, Amsterdam, 1976, pp. 121–172.)
2. P. J. Cohen, The independence of the Continuum Hypothesis. I. *Proc. Nat. Acad. Sci., U.S.A.* 50 (1963) 1143–1148, and II. *ibid.* 51 (1964) 105–110.
3. R. Dedekind, *Was sind und was sollen die Zahlen*?, F. Vieweg, Braunschweig, 1888. (English translation in *Essays on the Theory of Numbers*, Dover, New York, 1963, pp. 29–115.)
4. S. Feferman, Why a little bit goes a long way: Logical foundations of scientifically applicable mathematics, *PSA* 1992, *Vol. II* (1993), pp. 442–455.
5. S. Feferman, Gödel's program for new axioms: Why, where, how and what?, in (P. Hájek, ed.) *Gödel* '96, *Lecture Notes in Logic* 6 (1996), pp. 3–22.
6. S. Feferman and T. Strahm, The unfolding of non-finitist arithmetic (to appear).
7. H. Friedman, Finite functions and the necessary use of large cardinals, *Ann. of Math.* (2) (to appear).
8. H. Friedman, K. McAloon, and S. G. Simpson, A finite combinatorial principle which is equivalent to the 1-consistency of predicative analysis, in (G. Metakides, ed.) *Patras Logic Symposion*, North-Holland, Amsterdam, 1982, pp. 197–230.
9. H. Friedman, N. Robertson, and P. D. Seymour, The metamathematics of the graph minor theorem, in (S. G. Simpson, ed.) *Logic and Combinatorics*, Contemporary Mathematics 65, Amer. Math. Soc., Providence, 1987, pp. 229–261.
10. K. Gödel, Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I, *Monatsh. Math. Physik* 38 (1931), 173–198. (Reprinted with English translation in [14, pp. 144–195].)

11. K. Gödel, The consistency of the axiom of choice and of the generalized continuum hypothesis, *Proc. Nat. Acad. Sci., U.S.A.* 24 (1938), 556–557. (Reprinted in [**15**, pp. 26–27].)
12. K. Gödel, What is Cantor's continuum problem?, *Amer. Math. Monthly* 54 (1947), 515–525; errata 55, 151. (Reprinted in [**15**, pp. 176–187].)
13. K. Gödel, What is Cantor's continuum problem? (Revised and expanded version of [12], in (P. Benacerraf and H. Putnam, eds.) *Philosophy of Mathematics*: *Selected Readings*, Prentice-Hall, Englewood Cliffs, 1964, 258–273; reprinted in [15, pp. 254–270].)
14. K. Gödel, *Collected Works*, *Vol. I. Publications* 1929–1936 (S. Feferman, et al., eds.), Oxford University Press, New York, 1986.
15. K. Gödel, *Collected Works*, *Vol. II. Publications* 1938–1974 (S. Feferman, et al., eds.), Oxford University Press, New York, 1990.
16. W. Hanf, Incompactness in languages with infinitely long expressions, *Fund. Math.* 53 (1964), 309–324.
17. A. Kanamori, *The Higher Infinite*, Springer-Verlag, Berlin, 1994.
18. A. Kanamori and M. Magidor, The evolution of large cardinal axioms in set theory, in (G. H. Müller and D. S. Scott, eds.) *Higher Set Theory*, Lecture Notes in Mathematics 669 (1978), 99–275.
19. H. J. Keisler and A. Tarski, From accessible to inaccessible cardinals, *Fund. Math.* 53 (1964), 225–308; errata 57 (1965), 119.
20. A. Levy, Axiom schemata of strong infinity in axiomatic set theory, *Pacific J. Math.* 10 (1960), 223–238.
21. P. Maddy, Believing the axioms, I, *J. Symbolic Logic* 53 (1988), 481–511, and II, *ibid.*, 736–764.
22. D. A. Martin, Hilbert's first problem: The Continuum Hypothesis, in (F. Browder, ed.) *Mathematical Developments Arising from Hilbert Problems*, Proc. Symposia Pure Math. 28, Amer. Math. Soc., Providence, 1976, pp. 81–92.
23. J. Paris and L. Harrington, A mathematical incompleteness in Peano arithmetic, in (J. Barwise, ed.) *Handbook of Mathematical Logic*, North-Holland, Amsterdam, 1977, pp. 1133–1142.
24. G. Peano, *Arithmetices Principia, Nova Methodo Exposita*, Bocca, Turin (1889). (Reprinted in *Opera Scelte*, *Vol.* 2, Edizioni Cremonese, Rome, 1958, pp. 20–55, and in English translation in [28, pp. 83–97].)
25. D. S. Scott, Measurable cardinals and constructible sets, *Polish Acad. Sci. Math.* 9 (1961), 521–524.
26. S. G. Simpson, Nonprovability of certain combinatorial properties of finite trees, in (L. A. Harrington et al., eds.) *Harvey Friedman's Research on the Foundations of Mathematics*, North-Holland, Amsterdam, 1985, pp. 87–117.
27. A. Tarski, Some problems and results relevant to the foundations of set theory, in (E. Nagel, et al., eds.) *Logic, Methodology and Philosophy of Science. Proceedings of the* 1960 *International Congress*, Stanford University Press, Stanford, 1962, pp. 125–135.
28. J. van Heijenoort, *From Frege to Gödel. A Source Book in Mathematical Logic*, 1879–1931, Harvard University Press, Cambridge, 1967.
29. H. Weyl, *Das Kontinuum. Kritische Untersuchungen über die Grundlagen der Analysis*, Veit, Leipzig, 1918. (English translation: *The Continuum. A critical examination of the foundations of analysis*, University Press of America, Latham, 1987.)
30. H. Woodin, Large cardinal axioms and independence: The continuum problem revisited, *Math. Intelligencer* 16 (1994), 31–35.
31. E. Zermelo, Beweis, dass jede Menge wohlgeordnet werden kann, *Math. Ann.* 59 (1904), 514–516. (English translation in [28, pp. 139–141].)
32. E. Zermelo, Untersuchungen über die Grundlagen der Mengenlehre, I, *Math. Ann.* 65 (1908), 261–281. (English translation in [28, pp. 199–215].)
33. E. Zermelo, Über Grenzzahlen und Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre, *Fund. Math.* 16 (1930), 29–47.

SOLOMON FEFERMAN is professor of Mathematics and Philosophy, and Suppes Family Professor in Humanities and Sciences at Stanford University. He is past President of the Association for Symbolic Logic and is a Fellow of the American Academy of Arts and Sciences. He received his Ph.D. at U.C. Berkeley in 1957 under the direction of Alfred Tarski. He is Editor-in-Chief of the *Collected Works of Kurt Gödel*, and the author of numerous articles on mathematical logic, the foundations of mathematics, and the history of modern logic.
Stanford University, Stanford, CA 94305
sf@csli.stanford.edu

Statistical Independence and Normal Numbers: An Aftermath to Mark Kac's Carus Monograph

Gerald S. Goodman

1. INTRODUCTION. In 1958, Mark Kac delivered the prestigious Philips Lectures at Haverford College, which formed the basis for his Carus Monograph on Statistical Independence [**2**]. As a student, I had the privilege of attending those lectures. By foraging in their *aftermath*, meaning literally, "that which grows after the harvest," I have been led to the following ideas. They may be regarded as comprising a long footnote, or possibly a missing chapter, of that Monograph, composed in memory of its author by one of his erstwhile "guinea pigs."

In [**2**], Kac presents an arithmetical model for coin-tossing, due to Borel [**1**], which culminates in an analytical proof of the Strong Law of Large Numbers for Bernoulli trials. The idea is to identify the digits occurring in the binary expansion of a point $\omega \in \Omega = [0, 1)$,

$$\omega = \sum_{j=1}^{\infty} \frac{a_j(\omega)}{2^j}, \tag{1.1}$$

with the outcomes of a fair coin-tossing experiment. The j-th digit, $a_j(\omega)$, equals 1 if *tails* occurs on the j-th toss, and equals 0 if the outcome is *heads*. Then

$$S_n(\omega) = a_1(\omega) + a_2(\omega) + \cdots + a_n(\omega) \tag{1.2}$$

is the total number of tails in the first n tosses, and Borel's Strong Law of Large Numbers asserts that, when Ω is endowed with Lebesgue measure,

$$\frac{S_n(\omega)}{n} \to \frac{1}{2} \text{ a.e., as } n \to \infty. \tag{1.3}$$

Kac's proof of Borel's theorem makes use of the *Rademacher functions*. These are defined, for each $j \in \mathbb{N}$ and $\omega \in \Omega$, as

$$r_j(\omega) = 1 - 2a_j(\omega), \tag{1.4}$$

and they represent the net gain of a gambler who bets one florin on the outcome *heads* at the j-th toss. The orthonormality of the Rademacher functions on the unit interval Ω, and, indeed, of their finite products, is a consequence of the fact that the binary digits are statistically independent. It is this orthonormality that Kac exploits to give a simple proof of the Law of Large Numbers.

Kac goes on to interpret this result in terms of the existence of *simply normal numbers*. For such a number, the relative frequency of the 1's, among the first n digits in the binary expansion of its fractional part, tends to $1/2$ as $n \to \infty$. It follows from Borel's theorem that *almost all real numbers* have this property. Kac points out that a similar result holds for any integral base $b > 1$, sketches a proof, and concludes, with Borel, that almost all numbers are simply normal to every such base.

Now, it is known [**8**] that any number that is simply normal to all bases that are positive powers of an integral base b has the following property: Every string of finitely many b-adic digits occurs in the base b expansion of its fractional part with asymptotic relative frequency $1/b^m$, where m is the length of the string. Partial overlaps of the string with itself are counted as distinct occurrences. Numbers exhibiting this property are termed *normal to base b*, and the fact that almost every real number has this property is known as *The Normal Number Theorem* for b-adic digits.

Since Kac's use of orthonormality leads to such a simple proof of the case $m = 1, b = 2$, it is natural to ask whether his method can be adapted to give a direct proof of the Normal Number Theorem, at least for strings of binary digits. It turns out that this can be done by replacing Rademacher functions by their finite products, which themselves form an orthonormal system known as the *Walsh functions*. Kac presents them in an exercise, without any mention of their possible connection with normal numbers [**2**, p. 11].

That discovery was made by Mendès-France [**6**], who showed how base 2 normality can be characterized in terms of the Walsh functions. He then used this characterization to prove the Normal Number Theorem for binary digits. His approach, which, as he showed, generalizes to arbitrary integral bases $b > 1$, makes use of Haar measure, group characters, a generalized Weyl Criterion for asymptotically equidistributed sequences, and Birkhoff's Ergodic Theorem, applied to the dyadic map.

We shall show, instead, how the orthonormality of the Walsh functions leads directly to a simple proof, *a là* Kac, of the Normal Number Theorem for binary digits. The only additional idea required is the well-known observation, due to Wall [**10**] and used by Mendès-France, that a number is normal to base 2 if and only if the iterates, under the dyadic map, of its fractional part are uniformly distributed in the unit interval. Using this idea, our proof proceeds in the same spirit as one suggested by Kac—again, in the form of an exercise—for proving the classical theorem of Weyl [**11**] on the equidistribution of the fractional parts of multiples of irrational numbers [**2**, p. 41].

We can extend this approach to integral bases $b > 1$ by defining, with Mendès-France [**6**], the *b-adic Rademacher functions* as

$$r_j(\omega) = \exp\left(\frac{2\pi i\, b_j(\omega)}{b}\right), \text{ for all } \omega \in \Omega,\ j \in \mathbb{N},$$

where the b_j are the b-adic coefficients of the point ω,

$$\omega = \sum_{j=1}^{\infty} \frac{b_j(\omega)}{b^j},$$

and defining the b-adic Walsh functions as their power products. When $b = 2$, $b_j \in \{0, 1\}$ and Euler's formula ensures that the new definition of r_j agrees with the old one.

These b-adic Rademacher functions have mean zero, and all that is needed to establish their orthogonality (over the complex field) and further multiplicativity properties is the formula

$$\int_0^1 \prod_{j=1}^{k} \exp\big(i\mu_j b_j(\omega)\big)\, d\omega = \prod_{j=1}^{k} \int_0^1 \exp\big(i\mu_j b_j(\omega)\big)\, d\omega,$$

for suitable values of the real parameters μ_j and arbitrary values of $k \in \mathbb{N}$. But the validity of such a formula is guaranteed by the statistical independence of the

b_j's, as Kac well knew **[3]**, and its proof is virtually the same as one found in the opening pages of **[2]**.

In Section 8, we give the details, and show how the reasoning used to establish the base 2 normality of almost every real number goes through in the case of general integral bases b. Once done, it remains only to collect the exceptional sets in order to arrive at a new proof of The Normal Number Theorem, in its full force: *almost every real number is normal to every base.*

We then go on to examine more carefully the connection between the multiplicativity of the b-adic Rademacher functions, understood as the vanishing of their mixed moments, and the statistical independence of the b-adic coefficients. Using a device of Rényi's **[9]** we are able to draw a remarkable conclusion—the two notions are entirely equivalent! Since the multiplicativity property can be established easily by elementary analysis, this yields a new proof of the independence of the b-adic coefficients.

The same ideas can be applied to b-adic Walsh functions. While they themselves are not statistically independent, we find that the ones whose indices form a geometrical progression made up of fixed multiples of powers of b are. The statistical independence of the b-adic Rademacher functions, and thus of the b-adic coefficients, follows as a special case.

2. THE RADEMACHER FUNCTIONS. The functions r_j are defined by (1.4). Since the binary coefficients a_j satisfy

$$\int_0^1 a_j(\omega)\, d\omega = \frac{1}{2} \text{ for } j \in \mathbb{N},$$

it follows that

$$\int_0^1 r_j(\omega)\, d\omega = 0 \text{ for } j \in \mathbb{N}.$$

The statistical independence of the a_j implies that the r_j are also independent, and, since they each have mean zero, they satisfy the *Multiplicativity Formula*

$$\int_0^1 r_{j_1}(\omega) r_{j_2}(\omega) \cdots r_{j_k}(\omega)\, d\omega = 0,\ k \in \mathbb{N}, \tag{2.1}$$

whenever the subscripts are distinct. Since $r_j^2 = 1$ for every j, it follows from (2.1) that the Rademacher functions form an *orthonormal system* on Ω, that is,

$$\int_0^1 r_j(\omega) r_k(\omega)\, d\omega = \delta_{jk} \text{ for } j, k \in \mathbb{N}. \tag{2.2}$$

3. KAC'S PROOF [2], [3]. In view of (1.4), the limiting relation (1.3) is equivalent to the assertion that

$$\frac{R_n(\omega)}{n} \to 0 \text{ a.e., as } n \to \infty, \tag{3.1}$$

where

$$R_n(\omega) = r_1(\omega) + r_2(\omega) + \cdots + r_n(\omega) \tag{3.2}$$

is the difference between the number of heads and the number of tails in the first n tosses.

A direct computation using (2.1) and (2.2) shows that

$$\int_0^1 [R_n(\omega)]^4 \, d\omega = n + \frac{4!}{2!2!}\binom{n}{2} = n + 3n(n-1). \tag{3.3}$$

Consequently,

$$\sum_{n=1}^{\infty} \int_0^1 \left[\frac{R_n(\omega)}{n}\right]^4 d\omega < \infty,$$

and it follows from Beppo Levi's Theorem that

$$\sum_{n=1}^{\infty} \left[\frac{R_n(\omega)}{n}\right]^4 d\omega < \infty \text{ a.e.}$$

This implies that

$$\left[\frac{R_n(\omega)}{n}\right]^4 \to 0 \text{ a.e., as } n \to \infty,$$

which is clearly equivalent to (3.1).

4. A VARIANT OF KAC'S PROOF. The preceding proof uses (2.1) up to fourfold products, which is evidently a stronger property than the mere orthonormality of the Rademacher functions expressed by (2.2). However, a proof of (3.1) that uses only the orthonormality of the r_j (along with the uniform boundedness of their absolute values) can be based upon an argument employed by H. Weyl **[11]** in a similar context.

The orthonormality (2.2) implies that, for each $n \in \mathbb{N}$,

$$\int_0^1 [R_n(\omega)]^2 \, d\omega = n.$$

It follows that

$$\sum_{n=1}^{\infty} \int_0^1 \left[\frac{R_{n^2}(\omega)}{n^2}\right]^2 d\omega < \infty,$$

and, therefore, reasoning as before,

$$\frac{R_{n^2}(\omega)}{n^2} \to 0 \text{ a.e., as } n \to \infty. \tag{4.1}$$

Now, to each value of n, not a perfect square, there is a unique positive integer m_n such that $m_n^2 < n < (m_n + 1)^2$. Clearly, $m_n \to \infty$ as $n \to \infty$. In view of (3.2) and the fact that $|r_j| = 1$ for all j,

$$|R_n| \le |R_{m_n^2}| + |r_{m_n^2+1}| + \cdots + |r_n| \le |R_{m_n^2}| + 2m_n.$$

Dividing by m_n^2 and using (4.1) with n^2 replaced by m_n^2 yields

$$\frac{|R_n(\omega)|}{m_n^2} \to 0 \text{ a.e., as } n \to \infty.$$

Since $m_n^2 < n$, (3.1) follows.

5. SHIFT DYNAMICS. Define the *binary shift* (or *dyadic map*) T on Ω by the formula **[2**, p. 19 and p. 93]

$$T(\omega) = 2\omega \pmod 1. \tag{5.1}$$

Then, when ω is expressed in terms of its binary expansion (1.1), $T(\omega)$ takes the form

$$T(\omega) = \sum_{j=1}^{\infty} \frac{a_j \circ T(\omega)}{2^j} = \sum_{j=1}^{\infty} \frac{a_{j+1}(\omega)}{2^j}.$$

Making the convention that binary rationals have finite expansions, the uniqueness of the binary coefficients gives $a_{j+1}(\omega) = a_j \circ T(\omega)$ for $j \in \mathbb{N}$. Iteration yields

$$a_{j+k}(\omega) = a_j \circ T^k(\omega) \text{ for } j \in \mathbb{N},\ k \in \mathbb{N}_0, \tag{5.2}$$

and, in particular,

$$a_{1+k}(\omega) = a_1 \circ T^k(\omega) \text{ for } k \in \mathbb{N}_0. \tag{5.3}$$

Now, a_1 is the indicator function of the interval $[\frac{1}{2}, 1)$. In view of (5.3) and the definition (1.2) of S_n,

$$S_n(\omega) = a_1(\omega) + a_1 \circ T(\omega) + \cdots + a_1 \circ T^{n-1}(\omega).$$

Thus, $S_n(\omega)$ counts the number of times that the *orbit* of ω under T, that is,

$$\omega, T(\omega), \ldots, T^k(\omega), \ldots,$$

is found in $[\frac{1}{2}, 1)$ during the first n steps, starting from step 0. The Strong Law of Large Numbers can therefore be interpreted as saying that *the relative time that the orbit of* ω *occupies* $[1/2, 1)$ *tends to* $1/2$ *as* $n \to \infty$, for almost all starting points $\omega \in \Omega$, in accordance with [**2**, p. 93].

This dynamical interpretation of the Law of Large Numbers leads to a formulation of the Normal Number Theorem in an analogous way. Let

$$\alpha_1, \alpha_2, \ldots, \alpha_m \tag{5.4}$$

denote a string of binary digits of length m. Set

$$l = \sum_{j=1}^{m} \frac{\alpha_j}{2^{j-m}},$$

and let

$$I_l = \left[\frac{l}{2^m}, \frac{l+1}{2^m}\right). \tag{5.5}$$

Then I_l is the l-th *binary interval of order* m, and it consists of those points in Ω whose binary expansion starts out with (5.4). With this enumeration, the lexicographic order of the strings (5.4) is expressed by the natural order of the intervals (5.5), in terms of increasing values of l.

If we now denote by χ_{I_l} the indicator function of the interval I_l,

$$\frac{1}{n} \sum_{k=0}^{n-1} \chi_{I_l} \circ T^k(\omega)$$

is the average number of times the string (5.4) occurs among the first $n + m - 1$ binary coefficients of ω, where overlappings count as multiple occurrences, and an occurrence is marked at the moment when the string starts to appear. Any real number whose fractional part is ω will then be normal if, for each $m \in \mathbb{N}$,

$$\frac{1}{n} \sum_{k=0}^{n-1} \chi_{I_l} \circ T^k(\omega) \to \frac{1}{2^m}, \text{ as } n \to \infty, \tag{5.6}$$

where χ_{I_l} is the indicator function of a generic binary interval I_l of order m, and $1/2^m$ is its length. The Normal Number Theorem will, accordingly, be a consequence of the assertion that (5.6) holds a.e. on Ω for each such binary interval I_l.

6. THE WALSH FUNCTIONS. Kac introduces the Walsh functions, as follows. For each $j \in \mathbb{N}$, let

$$2j = 2^{j_1} + 2^{j_2} + \cdots + 2^{j_s}, \tag{6.1}$$

where $j_1 < j_2 < \cdots < j_s$ are in $\mathbb{N}$, be the unique expansion of the integer $2j$ in base 2. Then, for each $j \in \mathbb{N}$, define the *Walsh functions* by means of the formula

$$\omega_j(\omega) = r_{j_1}(\omega) r_{j_2}(\omega) \cdots r_{j_s}(\omega) \text{ for } \omega \in \Omega, \tag{6.2}$$

and set $w_0 \equiv 1$. The Walsh functions thus constitute an enumeration of all the finite products of Rademacher functions.

The Walsh functions, like the Rademacher functions, have a probabilistic interpretation in terms of coin tossing. For each $j \in \mathbb{N}$, w_j represents the net gain of the gambler who bets one florin on the outcome that the total number of *tails* occurring on the tosses $j_1, j_2, \ldots, j_s$ is *even.*

Because of (2.1), the Walsh functions are orthogonal on Ω, while the identity $|w_j^2| \equiv 1$ implies that they are orthonormal. For each $m \in \mathbb{N}$ the 2^m functions $w_0, w_1, \ldots, w_{2^m-1}$ are constant on binary intervals of order m. The range of any such function thus corresponds to a vector whose l-th component is the value taken by the function on the interval I_l. The orthogonality of the functions makes these vectors orthogonal and therefore linearly independent. It follows that the functions themselves are linearly independent. Consequently, any real-valued function f that is constant on binary intervals of order m can be written as a linear combination of the first 2^m Walsh functions:

$$f(\omega) = \sum_{j=0}^{2^m-1} \lambda_j w_j(\omega), \ \omega \in \Omega, \tag{6.3}$$

where the weights λ_j are the *Fourier-Walsh* coefficients of f:

$$\lambda_j = \int_0^1 f(\omega)\, \omega_j(\omega)\, d\omega, \ j = 0, \ldots, 2^m - 1.$$

In particular, the function χ_{I_l} can be written in such a form. In this case,

$$\lambda_0 = \int_0^1 \chi_{I_l}(\omega) w_0(\omega)\, d\omega = \frac{1}{2^m}, \tag{6.4}$$

while

$$\lambda_j = \int_0^1 \chi_{I_l}(\omega) w_j(\omega)\, d\omega = \pm \frac{1}{2^m}, \ j = 1, \ldots, 2^m - 1,$$

where the value of the $\pm$ sign is such as to make $\operatorname{sgn}[\lambda_j w_j] = +1$ on I_l.

7. PROOF OF THE NORMAL NUMBER THEOREM FOR BINARY DIGITS. In order to prove that (5.6) holds a.e. for any binary interval of the form (5.5) with $m \in \mathbb{N}$ arbitrary, observe that (1.4) and (5.2) imply that

$$r_{j+k}(\omega) = r_j \circ T^k(\omega) \text{ for all } \omega \in \Omega, j \in \mathbb{N}, k \in \mathbb{N}_0, \tag{7.1}$$

so that, in view of (6.1) and (6.2),

$$w_{2^k j}(\omega) = w_j \circ T^k(\omega) \text{ for all } \omega \in \Omega,\ j, k \in \mathbb{N}_0. \tag{7.2}$$

It is easy to verify that, for each $j \in \mathbb{N}$, the subset $w_j, w_{2j}, \ldots, w_{2^k j}, \ldots$ forms a multiplicative, orthonormal system, so the reasoning of the previous sections, applied to the sums $w_j(\omega) + w_{2j}(\omega) + \cdots + w_{2^{n-1}j}(\omega)$, instead of to the R_n, shows that, for each fixed $j \in \mathbb{N}$,

$$\frac{1}{n}\sum_{k=0}^{n-1} w_j \circ T^k(\omega) \to 0 \text{ a.e. as } n \to \infty,$$

while, trivially,

$$\frac{1}{n}\sum_{k=0}^{n-1} w_0 \circ T^k(\omega) \to 1 \text{ as } n \to \infty.$$

Consequently, using (6.3) with f replaced by χ_{I_l},

$$\begin{aligned}\frac{1}{n}\sum_{k=0}^{n-1} \chi_{I_l} \circ T^k(\omega) &= \frac{1}{n}\sum_{k=0}^{n-1}\sum_{j=0}^{2^m-1} \lambda_j w_j \circ T^k(\omega)\\ &= \sum_{j=0}^{2^m-1} \lambda_j\left(\frac{1}{n}\sum_{k=0}^{n-1} w_j \circ T^k(\omega)\right) \to \lambda_0 \text{ a.e. as } n \to \infty.\end{aligned}$$

Since, by (6.4), $\lambda_0 = 1/2^m$, (5.6) holds a.e., and the theorem is proved.

8. GENERALIZATION TO b-ADIC DIGITS. Suppose that $b > 1$ is any integral base. Let the b-adic expansion of a generic point $\omega \in \Omega$ be

$$\omega = \sum_{j=1}^{\infty} \frac{b_j(\omega)}{b^j}, \tag{8.1}$$

with the convention that b-adic rationals have finite expansions. Following **[6]**, define the *b-adic Rademacher functions*

$$r_j(\omega) = \exp\left(\frac{2\pi i\, b_j(\omega)}{b}\right) \text{ for all } \omega \in \Omega,\ j \in \mathbb{N}. \tag{8.2}$$

Thus, the r_j assume values in the cyclotomic group of b-th roots of unity, and each root is taken on a subset of Ω having measure $1/b$.

It follows at once that, for each $j \in \mathbb{N}$ and any integer d,

$$\int_0^1 r_j^d(\omega)\, d\omega = 0, \tag{8.3}$$

unless $d \equiv 0 \bmod b$. It is also evident that, when d is in the range from 0 to $b-1$,

$$\overline{r_j^d(\omega)} = r_j^{b-d}(\omega) \text{ for all } \omega \in \Omega,\ j \in \mathbb{N},$$

so the complex conjugates of powers of the r_j are expressible as positive powers of r_j. Accordingly, (8.3) holds also when r_j is replaced by its complex conjugate, as could have been seen directly.

As stated in the Introduction, we shall make use of the formula

$$\int_0^1 \prod_{j=1}^{k} \exp\big(i\mu_j b_j(\omega)\big)\, d\omega = \prod_{j=1}^{k} \int_0^1 \exp\big(i\mu_j b_j(\omega)\big)\, d\omega \tag{8.4}$$

for certain values of the real parameters μ_j and arbitrary $k \in \mathbb{N}$. The left-hand integral is one form of the multidimensional characteristic function (Fourier transform) of the first k b-adic coefficients, considered as random variables on Ω [**2**, p. 42], and it is their statistical independence that assures the equality of the two expressions. An elementary proof can be based upon Kac's demonstration of an analogous formula [**2**, p. 7].

Indeed, let $\mathbf{P}$ denote Lebesgue measure on Ω. Then

$$\int_0^1 \prod_{j=1}^{k} \exp\bigl(i\mu_j b_j(\omega)\bigr)\, d\omega$$

$$= \sum_{\beta_1,\ldots,\beta_k} \prod_{j=1}^{k} \exp(i\mu_j\beta_j)\mathbf{P}\{b_1(\omega) = \beta_1,\ldots, b_k(\omega) = \beta_k\},$$

where the β_j range independently over the set $0, 1, \ldots, b-1$. Using the statistical independence of the b_j's, the last expression becomes

$$\sum_{\beta_1,\ldots,\beta_k} \prod_{j=1}^{k} \exp(i\mu_j\beta_j) \prod_{j=1}^{k} \mathbf{P}\{b_j(\omega) = \beta_j\} = \sum_{\beta_1,\ldots,\beta_k} \prod_{j=1}^{k} \exp(i\mu_j\beta_j)\mathbf{P}\{b_j(\omega) = \beta_j\},$$

which, in turn, reduces to

$$\prod_{j=1}^{k} \sum_{\beta_j=0}^{b-1} \exp(i\mu_j\beta_j)\mathbf{P}\{b_j(\omega) = \beta_j\} = \prod_{j=1}^{n} \int_0^1 \exp\bigl(i\mu_j b_j(\omega)\bigr)\, d\omega,$$

as required.

Setting $\mu_j = 2\pi d_j/b$ in (8.4), where the d_j are integers between 0 and $b-1$, yields, in view of (8.3) and the defining equation (8.2), the *Multiplicativity Formula for the b-adic Rademacher functions*,

$$\int_0^1 r_1^{d_1}(\omega) r_2^{d_2}(\omega) \cdots r_k^{d_k}(\omega)\, d\omega = 0 \text{ for all } k \in \mathbb{N}, \tag{8.5}$$

unless all of the d's vanish, in which case the integral has the value 1. Recalling our comments about complex conjugates, we see that (8.5) continues to hold when any number of factors in the integrand are replaced by their conjugates. When $b = 2$, (2.1) can also be expressed in the present form, when k is suitably chosen and the exponents of the selected factors are set equal to 1, while the other exponents all vanish.

To generalize the Walsh functions, take any $j \in \mathbb{N}$ and make the partition

$$bj = d_{j_1} b^{j_1} + d_{j_2} b^{j_2} + \cdots + d_{j_s} b^{j_s}, \tag{8.6}$$

where the d's are integers in the range $1, \ldots, b-1$. Then set $w_0 \equiv 1$ and define the *b-adic Walsh functions*

$$w_j(\omega) = r_1^{d_{j_1}}(\omega) r_{j_2}^{d_{j_2}}(\omega) \cdots r_{j_s}^{d_{j_s}}(\omega),\ \omega \in \Omega,\ j \in \mathbb{N}. \tag{8.7}$$

This definition agrees with that of Mendès-France [**6**, p. 44].

By (8.2) and (8.7), $|w_j|^2 \equiv 1$ for all $j \in \mathbb{N}_0$. It then follows that the b-adic Walsh functions satisfy the orthonormality relations

$$\int_0^1 w_j(\omega)\overline{\omega_k(\omega)}\, d\omega = \delta_{jk} \text{ for } j, k \in \mathbb{N}_0, \tag{8.8}$$

for, by (8.5) and (8.7), the value of the integral can be 1 only if w_j and w_k are products of the same b-adic Rademacher functions, and their corresponding powers agree.

Instead of the dyadic map (5.1), consider the b-adic map $T(\omega) = b\omega \pmod 1$ and note that the uniqueness of the expansion (8.1) implies that $b_{j+k}(\omega) = b_j \circ T^k(\omega)$ for $j \in \mathbb{N}$, $k \in \mathbb{N}_0$. Thus, in view of (8.2), we also have

$$r_{j+k}(\omega) = r_j \circ T^k(\omega) \text{ for all } \omega \in \Omega,\ j \in \mathbb{N},\ k \in \mathbb{N}_0,$$

which is formally the same as (7.1). Consequently, by (8.7),

$$w_j \circ T^k(\omega) = r_{j_1+k}^{d_1}(\omega) r_{j_2+k}^{d_2}(\omega) \cdots r_{j_s+k}^{d_s}(\omega),\ \omega \in \Omega, \tag{8.9}$$

and so, in view of (8.7) and (8.6),

$$w_{b^k j}(\omega) = w_j \circ T^k(\omega) \text{ for all } \omega \in \Omega,\ j, k \in \mathbb{N}_0, \tag{8.10}$$

which generalizes (7.2).

Let m be any natural number, and take any string

$$\beta_1, \beta_2, \ldots, \beta_m \tag{8.11}$$

of m b-adic digits. Let I denote the b-adic interval of order m made up of those points in Ω whose b-adic expansion starts out with (8.11). Clearly, I has length $1/b^m$.

In view of (8.8), the first b^m b-adic Walsh functions, $w_0, w_1, \ldots, w_{b^m-1}$, are linearly independent over the complex numbers, and their span consists of all complex-valued functions that are constant on b-adic intervals of order m. In particular, χ_I can be written as

$$\chi_I(\omega) = \sum_{j=0}^{b^m-1} \lambda_j w_j(\omega),\ \omega \in \Omega, \tag{8.12}$$

where the weights λ_j are given by the formulas

$$\lambda_j = \int_0^1 \chi_I(\omega)\overline{w_j(\omega)}\, d\omega = \frac{\eta_j}{b^m},\ j = 1, \ldots, b^m - 1.$$

Here, η_j is a b-th root of unity for each j, and

$$\lambda_0 = \int_0^1 \chi_I(\omega)\overline{w_0(\omega)}\, d\omega = \frac{1}{b^m}. \tag{8.13}$$

It follows, again from (8.8), that, for each $j \in \mathbb{N}$, the functions

$$w_j, w_{bj}, \ldots, w_{b^k j}, \ldots \tag{8.14}$$

themselves form an orthonormal system on Ω. Accordingly, in view of (8.10), we can establish that, for each fixed $j \in \mathbb{N}$,

$$\frac{1}{n} \sum_{k=0}^{n-1} w_j \circ T^k(\omega) \to 0 \text{ a.e. as } n \to \infty, \tag{8.15}$$

by appealing to the reasoning of Section 4, provided we replace the second moments of the $R_n(\omega)$ there by the second moments of the *absolute values* of the sums

$$w_j(\omega) + w_{bj}(\omega) + \cdots + w_{b^{n-1}j}(\omega)$$

for $j = 1, \ldots, b^{m+1}$ and $n \in \mathbb{N}$. A more involved calculation, using the fact that, for each $k \in \mathbb{N}$,

$$w_{b^k j}(\omega) = r_{j_1+k}^{d_1}(\omega) r_{j_2+k}^{d_2}(\omega) \cdots r_{j_s+k}^{d_s}(\omega),$$

which follows from (8.9) and (8.10), and using, *mutatis mutandis*, the Multiplicativity Formula (8.5), shows that the functions (8.14) are multiplicative and that

$$\int_0^1 \left| w_j(\omega) + w_{bj}(\omega) + \cdots + w_{b^{n-1}j}(\omega) \right|^4 d\omega \leq n + 3n(n-1),$$

so Kac's original argument of Section 3 can still be used.

Once again,

$$\frac{1}{n} \sum_{k=0}^{n-1} w_0 \circ T^k(\omega) \to 1 \text{ as } n \to \infty.$$

Consequently, because of (8.12) and (8.15),

$$\frac{1}{n} \sum_{k=0}^{n-1} \chi_I \circ T^k(\omega) \to \lambda_0 \text{ a.e., as } n \to \infty,$$

just as in Section 7. But, in view of (8.13), $\lambda_0 = 1/b^m$. This proves that the relative frequency with which the string (8.11) occurs in the b-adic expansion of the fractional part of almost every real number tends to $1/b^m$ as $n \to \infty$, and that is enough to establish the Normal Number Theorem.

9. INDEPENDENCE AND MULTIPLICATIVITY IN BASE 2. Returning now to base 2, observe that

$$\frac{1 + \gamma\eta}{2} = \delta_{\gamma\eta}, \tag{9.1}$$

whenever $\gamma, \eta = \pm 1$, as noted by Rènyi [**9**, p. 130]. Consequently, for any $j \in \mathbb{N}$,

$$\frac{1 + \gamma_j r_j}{2} = \chi_{\{r_j = \gamma_j\}}, \tag{9.2}$$

where r_j is the j-th Rademacher function and $\gamma_j = \pm 1$. Since

$$\prod_{j=1}^{m} \chi_{\{r_j = \gamma_j\}} = \chi_{\{r_1 = \gamma_1, \ldots, r_m = \gamma_m\}},$$

it follows that for any $m \in \mathbb{N}$

$$\chi_{\{r_1 = \gamma_1, \ldots, r_m = \gamma_m\}} = \prod_{j=1}^{m} \left(\frac{1 + \gamma_j r_j}{2} \right). \tag{9.3}$$

Rényi uses this formula in the following way. He expands the product in (9.3) to get

$$\prod_{j=1}^{m} \left(\frac{1 + \gamma_j r_j}{2} \right) = \frac{1}{2^m} \left(1 + \sum_{j=1}^{m} \gamma_j r_j + \sum_{1 \leq j < l \leq m} \gamma_j \gamma_l r_j r_l + \cdots + \prod_{j=1}^{m} \gamma_j r_j \right). \tag{9.4}$$

Setting

$$\gamma_j = 1 - 2\alpha_j \text{ for } j = 1, \ldots, m, \text{ where } \alpha_j = 0 \text{ or } 1, \tag{9.5}$$

then gives, in view of (1.2) and (9.3), an expansion of the indicator function χ_{I_l} of a generic binary interval (5.5) of order m in terms of the first 2^m Walsh functions, as in our (6.3) *et seq.*, but without any recourse to orthonormality.

Rényi's approach can be used to motivate the introduction of finite products of Rademacher functions, and, therefore, of the Walsh functions, to deal with problems involving strings. However, it is also possible to employ these formulas in

another way to establish an unexpected connection between the multiplicativity of the Rademacher functions and their statistical independence.

Recall that, in Section 2, we used the statistical independence of the Rademacher functions, together with the vanishing of their means, to justify the Multiplicativity Formula (2.1). However, as noted by Kaczmarz and Steinhaus, this formula can also be established directly by calculus, without appealing to the notion of statistical independence [**4**, p. 125*f*].

To do so, note first that (7.1) implies that

$$r_j = r_1 \circ T^{j-1} \text{ for } j \in \mathbb{N},$$

where T is the binary shift given by (5.1). For fixed j, let l vary in the range from 0 to $2^{j-1} - 1$, and denote by I_l the l-th dyadic interval of order $j - 1$. Since T^{j-1} is linear on each such I_l and maps it onto Ω, while the derivative of T^{j-1} on I_l is 2^{j-1}, the change-of-variable formula gives

$$\int_{I_l} r_j(\omega)\, d\omega = \int_{I_l} r_1 \circ T^{j-1}(\omega)\, d\omega = \frac{1}{2^{j-1}} \int_0^1 r_1(\omega)\, d\omega = 0 \tag{9.6}$$

for each I_l. Summing on l then yields

$$\int_0^1 r_j(\omega)\, d\omega = 0 \text{ for } j \in \mathbb{N}, \tag{9.7}$$

as asserted at the start of Section 2.

To arrive at the Multiplicativity Formula (2.1), we can now employ an argument used by Kac in a different connection [**2**, p. 27]. Order the subscripts so that j_k is the largest, and then write

$$\int_0^1 r_{j_1}(\omega) r_{j_2}(\omega) \cdots r_{j_k}(\omega)\, d\omega = \sum_{l=0}^{2^{j_k-1}-1} \int_{I_l} r_{j_1}(\omega) r_{j_2}(\omega) \cdots r_{j_k}(\omega)\, d\omega,$$

where the I_l are the dyadic intervals of order $j_k - 1$. Since the first $k - 1$ functions $r_{j_1}, r_{j_2}, \ldots, r_{j_{k-1}}$ are *constant* on each such interval, while, by (9.6) with $j = j_k$, the integral of r_{j_k} over each one is zero, the integrals in the sum at right all vanish, and the Multiplicativity Formula follows.

With this result in hand, we can establish the statistical independence of the Rademacher functions by proceeding in the following way. We integrate (9.2) over Ω and use (9.7) to find that for every $j \in \mathbb{N}$

$$\mathbf{P}\{r_j = \gamma_j\} = \frac{1}{2} \text{ whenever } \gamma_j = \pm 1. \tag{9.8}$$

Integrating (9.3) over Ω gives

$$\mathbf{P}\{r_1 = \gamma_1, \ldots, r_m = \gamma_m\} = \int_0^1 \prod_{j=1}^{m} \left(\frac{1 + \gamma_j r_j}{2} \right).$$

Now integrating (9.4) and making use of the Multiplicativity Formula along with (9.8) gives

$$\mathbf{P}\{r_1 = \gamma_1, \ldots, r_m = \gamma_m\} = \frac{1}{2^m} = \prod_{j=1}^{m} \mathbf{P}\{r_j = \gamma_j\}. \tag{9.9}$$

Since $m \in \mathbb{N}$ is arbitrary, it follows that *multiplicativity of the Rademacher functions implies their statistical independence*.

The preceding reasoning is not limited to the case of Rademacher functions. As a further application, recall from Section 7 that the Walsh functions

$$w_j, w_{2j}, \dots, w_{2^k j}, \dots \tag{9.10}$$

are, for fixed $j \in \mathbb{N}$ and varying $k \in \mathbb{N}$, multiplicative. When $j = 1$, these functions reduce to the Rademacher functions, as is evident from (6.1) and (6.2). Noting that they have mean zero, we can repeat our argument and conclude that the functions (9.10) are, in fact, statistically independent for *every* fixed $j \in \mathbb{N}$.

If, finally, we make the substitution (9.5) in (9.9), and do the same in (9.8), then (9.9) becomes

$$\mathbf{P}\{a_1 = \alpha_1, \dots, a_m = \alpha_m\} = \frac{1}{2^m} = \prod_{j=1}^{m} \mathbf{P}\{a_j = \alpha_j\}.$$

Thus, *statistical independence of the binary digits is, itself, a direct consequence of multiplicativity of the Rademacher functions, and, therefore, the two notions are entirely equivalent.*

10. EXTENSION TO GENERAL BASES b. All the reasoning of the preceding section can be carried over to integral bases $b > 1$, once we find a suitable generalization of Rényi's identity (9.1). To this end, let γ and η be b-th roots of unity and consider the expression

$$\sum_{d=0}^{b-1} \gamma^{b-d} \eta^d = \sum_{d=0}^{b-1} \left(\frac{\eta}{\gamma}\right)^d.$$

Since $\gamma^b = \eta^b = 1$, the quotient η/γ is also a b-th root of unity, and, therefore, the right-hand sum vanishes unless $\gamma = \eta$. Consequently,

$$\frac{1 + \sum_{d=1}^{b-1} \gamma^{b-d} \eta^d}{b} = \gamma_{\gamma\eta},$$

and we have found the identity we need.

Proceeding as before, for each $j \in \mathbb{N}$ we get

$$\frac{1 + \sum_{d=1}^{b-1} \gamma_j^{b-d} r_j^d}{b} = \chi_{\{r_j = \gamma_j\}}, \tag{10.1}$$

and, since, by (8.3),

$$\int_0^1 r_j^d(\omega)\, d\omega = 0 \text{ for } j \in \mathbb{N} \text{ and } d = 1, \dots, b-1,$$

integrating (10.1) gives

$$\int_0^1 \chi_{\{r_j = \gamma_j\}} = \mathbf{P}\{r_j = \gamma_j\} = \frac{1}{b}. \tag{10.2}$$

This time,

$$\chi_{\{r_1 = \gamma_1, \dots, r_m = \gamma_m\}} = \prod_{j=1}^{m} \chi_{\{r_j = \gamma_j\}} = \prod_{j=1}^{m} \left(\frac{1 + \sum_{d=1}^{b-1} \gamma_j^{b-d} r_j^d}{b} \right). \tag{10.3}$$

Expanding the second product, we get an expression of the form

$$\frac{1}{b^m}(1 + \text{sums of weighted power-products of the } r_j), \tag{10.4}$$

where the weight assigned to each r_j^d is γ_j^{b-d}. Thus, integrating (10.3) over Ω and using the Multiplicativity Formula (8.5) for the b-adic Rademacher functions, yields, in view of (10.2),

$$\mathbf{P}\{r_1 = \gamma_1, \ldots, r_m = \gamma_m\} = \frac{1}{b^m} = \prod_{j=1}^{m} \mathbf{P}\{r_j = \gamma_j\}. \tag{10.5}$$

Accordingly, *multiplicativity of the b-adic Rademacher functions implies their statistical independence*, just as in the *case* $b = 2$.

In Section 8, we introduced subsets of b-adic Walsh functions of the form

$$w_j, w_{bj}, \ldots, w_{b^k j}, \ldots, \tag{10.6}$$

for fixed values of $j \in \mathbb{N}$ and varying $k \in \mathbb{N}$, and pointed out that they were multiplicative, in the sense that

$$\int_0^1 w_{b^{j_1} j}(\omega) w_{b^{j_2} j}(\omega) \cdots w_{b^{j_k} j}(\omega)\, d\omega = 0$$

whenever the exponents are distinct. We can now show that they are statistically independent for each $j \in \mathbb{N}$ by extending this formula in the following way. Let b' be the smallest positive integer such that $w_j^{b'} \equiv 1$. Then, as a consequence of (8.5) and the choice of b', the functions (10.6) are multiplicative in the extended sense that

$$\int_0^1 w_j^{d_1'}(\omega) w_{bj}^{d_2'}(\omega) \cdots w_{b^k j}^{b_k'}(\omega)\, d\omega = 0$$

for every $k \in \mathbb{N}$, where the d's are restricted to the values $0, 1, \ldots, b' - 1$ and do not all vanish. It is thus possible to apply the foregoing reasoning to them, by replacing b by b' in the formulas above. This yields the statistical independence of the functions (10.6), and it generalizes (10.5), which, in view of (8.6) and (8.7), corresponds to the case $j = 1$.

Returning now to (10.3) and setting

$$\gamma_j = \exp\left(\frac{2\pi i \beta_j}{b}\right) \tag{10.7}$$

for $j = 1, \ldots, m$, with $\beta_j \in \{0, 1, \ldots, b-1\}$, we get, in view of (8.2) and the expansion (10.4), a formula for the indicator functions

$$\chi_{\{b_1 = \beta_1, \ldots, b_m = \beta_m\}}$$

of b-adic intervals of order m, in terms of weighted power-products of the r_j. The resulting formula is, clearly, comparable to our (8.5), *et seq.*, and it can be used to motivate the introduction of power-products of the b-adic Rademacher functions, and, thus, of the b-adic Walsh functions, in order to deal with problems involving b-adic strings.

It can also be employed to establish an equivalence between multiplicativity of the b-adic Rademacher functions and the statistical independence of the b-adic digits, generalizing the case $b = 2$ treated in the previous section. Indeed, it is

enough to make the substitution (10.7) in (10.5) to get

$$\mathbf{P}\{b_1 = \beta_1, \ldots, b_m = \beta_m\} = \frac{1}{b^m} = \prod_{j=1}^{m} \mathbf{P}\{b_j = \beta_j\}.$$

Consequently, *multiplicativity of the b-adic Rademacher functions implies statistical independence of the b-adic digits.*

Since there is no difficulty in extending to a general base b the argument used in the previous section to establish the Multiplicativity Formula for Rademacher functions, without having recourse to their statistical independence, we conclude that *statistical independence of the b-adic digits is equivalent to the multiplicativity property* (8.5) *possessed by the b-adic Rademacher functions.*

One way to look at this result is as follows. Recall that, in Section 8, we derived the Multiplicativity Formula (8.5) from the formula

$$\int_0^1 \prod_{j=1}^{k} \exp(i\mu_j b_j(\omega))\, d\omega = \prod_{j=1}^{k} \int_0^1 \exp(i\mu_j b_j(\omega))\, d\omega, \tag{8.4}$$

by setting $\mu_j = 2\pi d_j/b$, where the d_j are integers in the range from 0 to $b-1$ and j varies from 1 to k. As shown there, (8.4) is a direct consequence of the statistical independence of the functions b_j.

However, it is known that the validity of (8.4) for *all* real values of the parameters μ_j *implies* statistical independence of the otherwise quite arbitrary real functions b_j, and Kac himself was among the first to prove it, by use of Fourier analysis [**3**].

In effect, what we have found is a refinement of Kac's sufficiency result, in a more algebraic setting, where the additive group of the reals has been replaced by the group of integers mod b. For, in restricting the b_j's to take integral values in the range from 0 to $b-1$, instead of arbitrary real values, we have established that the validity of (8.4) for the *narrow* range of values of μ_j indicated above is already enough to ensure statistical independence of the b_j's.

The argument we have used cannot be regarded as new, even though its application to Rademacher and Walsh functions is. It is a reduction to the case of the cyclotomic group of a method suggested by Rényi [**9**, p. 171] for proving a more general result about the statistical independence of families of random variables, attributed to Kantorovic. Rényi would, no doubt, have appreciated that the method he proposed could be applied, trivially, to the classical Rademacher functions in order to establish their statistical independence, but he seems to have been unaware, as was Kac, of Mendès-France's extension of the Rademacher functions to the b-adic case and, thus, missed the opportunity to apply his method there.

The reader interested in learning more about normal numbers can consult Niven's Carus Monograph [**7**] and the treatise by Kuipers and Niederreiter [**5**], which contains an extensive bibliography.

ACKNOWLEDGMENT. This work was supported in part by the Consiglio Nazionale delle Ricerche and the Italian Ministry of the Universities and Scientific and Technological Research.

REFERENCES

1. E. Borel, Sur les probabilités dénombrables e leurs applications arithmétiques, *Rend. Circ. Mat. Palermo* **27** (1909) 247–271.
2. M. Kac, *Statistical Independence in Probability, Analysis and Number Theory*, Carus Math. Monographs, no. 12, Wiley, New York 1959.

3. M. Kac, Sur les fonctions indépendentes, I, *Studia Math.* **6** (1936) 46–58.
4. S. Kaczmarz, and H. Steinhaus, *Theorie der Orthogonalreihen*, Monografie Matematyczne, vol. 6, Warsaw and Lwow, 1935.
5. L. Kuipers, and H. Niederreiter, *Uniform Distribution of Sequences*, Pure and Applied Math., no. 29, Wiley, New York, 1974.
6. M. Mendès-France, Nombres normaux. Applications aux fonctions pseudo-aléatoires, *J. d'Anal. Math.* **20** (1967) 1–56.
7. I. Niven, *Irrational Numbers*, Carus Math. Monographs, no. 11, Wiley, New York, 1956.
8. S. S. Pillai, On normal numbers, *Proc. Indian Acad. Sci.*, Sect. A **12** (1940) 179–184.
9. A. Rényi, *Foundations of Probability*, Holden-Day, San Francisco, 1970.
10. D. D. Wall, *On Normal Numbers*, Ph.D. Thesis, Univ. of California, Berkeley, 1949.
11. H. Weyl, Über die Gleichverteilung von Zahlen mod. Eins, *Math. Ann.* **77** (1916) 313–352.

GERALD S. GOODMAN took degrees in mathematics at Haverford, Bryn Mawr, and Stanford, where he was a pupil of Loewner. His doctoral thesis was devoted to Loewner's theory of transformation semigroups, applied to conformal mapping. A chance encounter with Rényi led him to use those ideas to obtain new results concerning Markov chains. Later on, he applied probabilistic tools to Loewnerian semigroups, such as totally positive matrices. Although he has published papers in real analysis, control theory, and fractal image generation, some consider his finest mathematical achievement was finding a permanent post in a Renaissance city.
Statistics Department, University of Florence, Viale Morgagni, 59, 50134 Florence, ITALY
goodman@ds.unifi.it

...it did not take the Miloslavskys long to find a pretext for arresting Matveev. That educated gentleman was foolish enough to be found with a book of algebra in his baggage, which was, naturally, taken to be a form of black magic.

Even Nikita, when he heard of the arrest of his mentor, could only shake his head and remark: "He was asking for trouble. What did he want with such stuff anyway?"

Edward Rutherfurd, *Russka, The Novel of Russia*, Crown Publishers, Inc., New York, 1991, pp. 347–348.

Contributed by George Dorner, William Rainey Harper College, Palatine, IL

The Velocity Dependence of Aerodynamic Drag: A Primer for Mathematicians

Lyle N. Long and Howard Weiss

1. INTRODUCTION. If you pick up nearly any elementary ordinary differential equations text or calculus text, you are likely to find a short section, or at least a problem, on the motion of a body subject to some type of drag force along with a calculation of the body's terminal velocity. Two favorite examples seem to be the motion of a projectile like a baseball and the motion of a skydiver/parachutist, both through the air. By a *skydiver* we mean a person falling without his parachute open. Most textbook authors model the motion of these objects using a drag force that depends linearly in the velocity. Unfortunately, the physical assumption about the linear dependence of the drag force on velocity is often incorrect, and thus the model's predictions are physically implausible.

In particular it was surprising to see the faulty linear resistance model for a parachutist's velocity used in the popular calculus reform text by Hughes-Hallet, Gleason, et al. [**9**, p. 515], since the reform movement prides itself on concern for realistic applications. The first edition of this text even supplied non-referenced *observed data* to fit its linear model. The authors state "The fact that there is good agreement between the observed and predicted data suggest that our assumption about the air resistance is reasonable." The recent second edition omits the table of *observed data* but not the flawed model.

The purpose of this note is to explain the dependence of the drag force on velocity for a general mathematical audience and to present a few realistic models. Section 5 contains an interesting model (with a closed form solution) for re-entry of the space shuttle into the earth's atmosphere.

Dimensional analysis is an important tool in aerodynamics and fluid dynamics, and can be used to obtain key results (5) and (6). To help give mathematicians some insight into the spirit of this important technique, we present in Section 6 an amusing application of dimensional analysis to prove the Pythagorean Theorem.

The science of modeling drag is more physical and empirical than mathematical, and it relies on the results of many wind-tunnel experiments. There has been a significant amount of *theoretical* work in the engineering literature, but few of the results can be considered completely rigorous by mathematical standards. There are also large gaps in our understanding of basic properties of the Navier Stokes equation. In particular, there are important unsolved problems on the large-time existence and uniqueness of solutions of the Navier-Stokes equation in three dimensions.

For detailed information on the aerodynamics and fluid mechanics pertinent to this paper, see [**7**], [**8**], [**11**], [**12**], [**19**], and [**22**].

2. THE BASIC EQUATIONS OF MOTION. Any body moving through a fluid such as water or air creates a drag force that tends to retard its motion. Such motion is usually described by the Navier-Stokes (nonlinear partial differential) equations. In elementary textbooks, the motion is always assumed to be one

dimensional, e.g., the ball is dropped and the skydiver has no horizontal movement and there is no wind. We observe in Section 4 that this assumption does not permit modeling of a modern parachute. Many (if not most) elementary mathematics textbook authors assume that the drag force for a baseball or skydiver/parachutist moving in air is proportional to the velocity v of the falling body, and at least one leading freshman physics text makes this assumption. This leads to the linear differential equation of motion

$$m\frac{dv}{dt} = mg - k_1 v, \tag{1}$$

where k_1 is a constant (whose physical meaning is rarely discussed), m is the mass of the body, and g is the gravitational constant. This linear differential equation can be solved easily to obtain the body's velocity as a function of time, beginning at rest:

$$v_1(t) = \frac{mg}{k_1}(1 - \exp(-k_1\, t/m)), \quad v_1(0) = 0. \tag{2}$$

The terminal velocity is $\lim_{t\to\infty} v_1(t) = mg/k_1$. This terminal velocity is just the equilibrium solution of (1) and could have been obtained easily directly from (1) without explicitly solving the differential equation since physically, the terminal velocity corresponds to the motion when the drag force *precisely equals* the weight mg of the falling object. Any such simple model is necessarily a great simplification of the Navier-Stokes equations for the actual motion of a ball or skydiver. While this model may be correct for bodies that are falling in a vat of heavy oil or for tiny particles of dust or aerosol in air, it is *grossly incorrect* for large bodies falling in air.

Calculations predict and experiments confirm that in air, the drag force on a ball or a skydiver/parachutist can be well *approximated* by a force that is proportional to the *square* of the velocity v^2 (and *not* to the velocity v). The v^2 model for the drag force leads to the nonlinear equation of motion

$$m\frac{dv}{dt} = mg - k_2 v^2, \tag{3}$$

where k_2 is a constant. This is a separable equation, which can be solved easily to obtain the body's velocity as a function of time, beginning at rest:

$$v_2(t) = \sqrt{\frac{mg}{k_2}}\tanh\left(t\sqrt{\frac{k_2 g}{m}}\right), \quad v_2(0) = 0. \tag{4}$$

The terminal velocity is $\lim_{t\to\infty} v_2(t) = \sqrt{mg/k_2}$, which is just the equilibrium solution of (3) and could have been obtained easily directly from (3).

Table 1 contains the experimentally determined terminal velocities for various objects moving through the air. There is a wide range of values for the terminal velocity of a skydiver because the terminal velocity strongly depends on his body position and is considerably higher (almost by a factor of two) during a head first *nose dive* or *feet first* dive than during a fall in the *spread eagle belly-to-Earth* position. The former positions are highly unstable and are difficult to maintain for more than a few seconds. In order to minimize the strong shock to the body at deployment, beginners typically reduce their free fall speed to about 50 m/s before deploying their parachute.

TABLE 1. Approximate terminal velocities for various objects (from Table 9.1 in [**4**])

Object	Weight (kg)	Terminal Velocity (m/s)
iron ball (shot)	7.3	145
Skydiver	72.6 + 19 (equipment) = 91.6	45 to 80 +
Football	0.41	45
Baseball	0.15	42
Golfball	0.05	40
Softball	0.18	80
Tennis ball	0.06	36
Basketball	0.6	20
Ping-Pong ball	0.003	9
Parachutist (round canopy)	72.6 + 19 (equipment) = 91.6	5

3. SMALL AND LARGE REYNOLDS NUMBERS FLOWS. In general, the drag force depends on many factors including the density and viscosity of the fluid, and the geometry, surface material, surface regularity, and velocity of the body. The dimensionless *Reynolds number* of the fluid plays a key role in determining the drag force, and is defined by

$$R = \frac{\rho d v}{\mu}, \quad \text{or} \quad R = \frac{dv}{\nu},$$

where ρ is the density of the fluid, v is the velocity of the body in the fluid, μ is the viscosity of the fluid, $\nu = \mu/\rho$, and d is a characteristic length (see Table 2). This characteristic length could be a radius, a diameter, a chord length, a body length, etc. depending on what aspect of the problem one is studying. Note that a slowly moving object may have a large Reynolds number if the object is large or the viscosity is small. Turbulent flows are typically associated with large Reynolds numbers, while laminar flows are typically associated with small Reynolds numbers.

TABLE 2. Typical Reynolds numbers for various objects moving in air

Object	Characteristic Length	Typical Reynolds Number
Submarine	Length	300,000,000
Small aircraft	Chord	5,000,000
Parachutist	Diameter	2,500,000
Skydiver	Diameter	1,000,000
Baseball	Diameter	250,000
Model airplane	Chord	50,000
Butterfly	Chord	7,000
Dust particle	Diameter	1

If the Reynolds number is *small*, meaning $R \ll 1$, the Navier-Stokes equation is considerably simplified and the equation of motion reduces to a linear partial differential equation. Strictly speaking, one should assume $R \ll 1$, but the approximation is often reasonable for $R \approx 1$. Stokes analyzed this linear differential equation and found the following formula for the drag force, F_D, on a sphere of radius r moving in the fluid [**21**, p. 217]:

$$F_D = 6\pi\mu r v. \tag{5}$$

This expression is exact in the limit as the Reynolds number goes to zero. Thus, the drag force is proportional to the velocity and the radius of the sphere. Since

the fluid density does not appear in the linear partial differential equation, the form of formula (5) can also be obtained with simple dimensional arguments: if the drag force depends only on μ, r, and v, one shows the only function of these quantities that has the units of force is $F_D = C\mu r v$, where C is a constant. A rigorous argument can be based on the Pi theorem of Vaschy and Buckingham [**1**, p. 42], [**2**], [**3**].

Formula (5) can be extended to flows with non-zero Reynolds number. Using techniques in the theory of matched asymptotic expansions, the Stokes approximation (5) can be improved [**17**] to an asymptotic expansion of the form

$$F_D = 6\pi\mu r v\left(1 + \frac{3}{8}R + \frac{9}{40}R^2 \log R + O(R^2)\right).$$

Table 3 contains the values of μ and ν for oil, water, and dry air at 100° F. It is known that the viscosity of oils increases rapidly with decreasing temperature.

TABLE 3. Typical values of μ and ν at 100 degrees F

Medium	μ (kg/m sec)	ν (m^2/ sec)
castor oil	0.29	2.8×10^{-4}
water	0.686×10^{-3}	0.691×10^{-6}
dry air	0.19×10^{-4}	1.9×10^{-5}

It has been determined experimentally that (5) is valid for Reynolds numbers $R < 1$ and that a similar dependence occurs in this range of R for bodies with other shapes, i.e., the drag force

$$F_D \approx \text{constant} \times \mu v,$$

where the constant is independent of v. This can be rewritten as $F_D = kv$, where $k = \text{constant} \times \mu$ (see (1) and (2)). From Table 2 we see that baseballs and skydivers/parachutists have $R \gg 1$.

There are some interesting implications of low Reynolds number flows in biology. In particular, [**20**] describes the role of terminal velocity in pollen dispersal, while [**6**] answers the question "Why are there so few aerial plankton?" by explaining how high atmospheric terminal velocities confound the ability of turbulence to keep organisms afloat.

Although there are interesting flows where the drag depends linearly on velocity, they are typically associated with small objects such as raindrops, dust particles, etc. The book [**14**] contains a discussion of modeling falling raindrops over a wide range of Reynolds numbers.

When the Reynolds number is large, but not *too* large, the flow may remain laminar. These cases can be studied using the Navier-Stokes equations in the thin boundary layer around the body where this flow is assumed to be laminar. The resulting equations are called Prandtl's equations [**11**] and one can conclude that (at least for a certain range of R) the drag force is independent of the viscosity. One then uses facts about the Bernoulli equation or dimensional analysis to conclude that

$$F_D \approx \text{constant} \times \rho A v^2, \tag{6}$$

where the constant depends only on the shape and surface characteristics of the body. Numerous experiments in wind tunnels and in aircraft flight tests during the past 80 years have verified that this formula is valid for Reynolds numbers between 3×10^2 and 3×10^5.

For flows in the range of Reynolds numbers, it is customary to introduce the *drag coefficient*, C_D, which is the dimensionless quantity defined by

$$C_D \equiv \frac{F_D}{\frac{1}{2}\rho A v^2}. \tag{7}$$

With this definition and (6), we have C_D = constant, i.e., the drag coefficient depends only on the shape and surface characteristics of the body and the Reynolds number. Thus in (3) and (4), the constant $k = \frac{1}{2}C_D \rho A$. Furthermore, the *dynamic pressure* $q \equiv \frac{1}{2}\rho v^2$ plays a fundamental role in aerodynamic theory **[7]**. For instance, when the space shuttle Challenger exploded, it was operating in a high dynamic pressure regime. Very fast aircraft need to operate at high altitude (where ρ is relatively small) to avoid excessive dynamic pressure and catastrophic damage to the aircraft.

For smooth spheres having Reynolds numbers in the range 10^3 to 3×10^5, the drag coefficient is approximately 0.47, while for Reynolds numbers greater than 3×10^5, the drag coefficient is approximately 0.20 (see Figure 1). The text **[22]** contains a good exposition of *sphere drag* for R between 1 and 10^6.

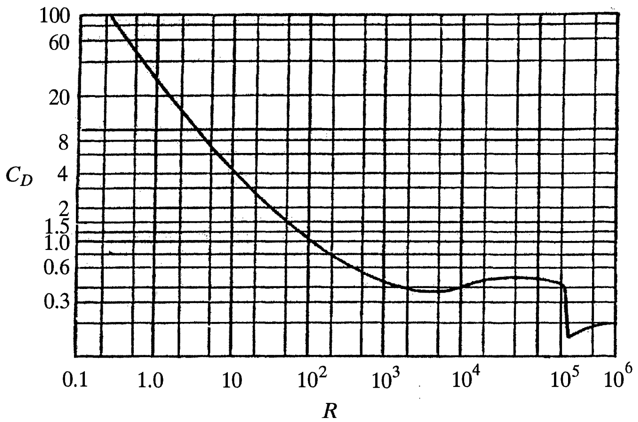


Figure 1. Drag coefficient C_D for a sphere as a function of Reynolds number R (from Figure 34 in **[11]**)

It follows from (4) and (7) that the terminal velocity for a sphere falling in air is approximately

$$v_T = \sqrt{\frac{2W}{\rho C_D \pi r^2}},$$

where W is the weight of the sphere, ρ is the density of air at sea level, and r is the radius of the sphere. The density ρ is a complicated function of temperature, humidity, and pressure (which varies with altitude) so this equation is only an approximation.

The motion of a baseball, with its rough surface, is actually considerably more complicated to model accurately than the motion of a smooth sphere **[13]**.

4. SKYDIVING AND PARACHUTING. We now discuss the motion of a skydiver and a parachutist; useful technical references are **[8]**, **[10]**, **[15]**, and **[16]**. Just as for a sphere falling in air, the terminal velocity for a skydiver is approximately

$$v_T = \sqrt{\frac{2W}{\rho C_D A}}, \tag{8}$$

where W is the combined weight of the skydiver and parachute, A is the effective cross-sectional area of the skydiver, and the density of air is $\rho = 1.225\,\mathrm{kg/m^3}$. Solving for the drag coefficient C_D we obtain

$$C_D = \frac{2W}{\rho A v_T^2} = \frac{W}{qA}, \tag{9}$$

where q is the dynamic pressure corresponding to terminal velocity.

If a 72.6 kg skydiver carrying a 19 kg load (91.6 kg = 867 N) attains a terminal velocity of 49 m/s (in the belly-to-earth position) and has a cross-sectional area of 0.56 $\mathrm{m^2}$, it follows from (9) that $C_D \approx (2 \times 867)/(1.225 \times 0.56 \times 49^2) = 1.05$.

Moreover, if our skydiver attains a terminal velocity of 67 m/s (in the head down or feet down position) and has a cross-sectional area of 0.2 $\mathrm{m^2}$ in this position, it again follows from (9) that $C_D \approx (2 \times 867)/(1.225 \times 0.2 \times 67^2) = 1.57$. Actually, even if the skydiver could maintain the head down or feet down position over a long period of time, his rate of descent would continually slow due to the increasing density of air at lower altitudes.

In the 1960s, a 72.6 kg beginner sport parachutist might have used a circular parachute with a canopy area of 74.8 $\mathrm{m^2}$, and would have carried about a 22.7 kg load (95.3 kg = 934 N) **[15]**. The parachute would have had $C_D \approx 0.8$. It follows from (8) that the terminal velocity for the parachutist is approximately $[(2 \times 934)/(1.225 \times 0.8 \times 74.7)]^{1/2} = 5.1$ m/s. Many measurements have confirmed that this prediction is quite a close approximation to the actual terminal velocity.

The sport parachutes used today bear little resemblance to the old classical round canopies, although the latter are still preferred by the military. The military's round canopies also have a relatively small area, which results in much harder landings than with modern sport canopies. Today, nearly all jumpers use either *square* (actually rectangular) or elliptical canopies, made from a non-porous material. When open, these canopies act like an airplane wing or an airfoil, and generate lift throughout the flight; they do not work by drag alone and are more like gliders than umbrellas. In addition, these modern square or elliptical canopies actually have *brakes* that the parachutist can apply close to the ground to achieve a gentle landing. Because of the lift that these canopies generate, their motion can not be modeled solely by the simple v^2 drag force model with the force parallel to motion.

However, we can obtain a reasonable model of a modern parachute by adding an extra term to (3) corresponding to the lift generated by the canopy. These are the same equations that are used to model flight of an unpowered airplane (a glider) or re-entry of the space shuttle into the earth's atmosphere. The force due to lift, F_L, is proportional to the square of the velocity, but now it is important to consider the horizontal component of motion–thus the new model is necessarily two dimensional and (3) is replaced by a pair of coupled nonlinear equations **[7]**.

It is convenient to work in a special (rotating) coordinate system centered on the center of the earth. Letting V denote the tangential component of velocity for the unpowered aircraft, the equations of motion are

$$m\frac{dV}{dt} = -F_D - W\sin\theta, \qquad m\frac{V^2}{r_E} = -F_L + W\cos\theta \tag{10}$$

where θ denotes the *climb angle*, r_E is the radial distance of the aircraft to the center of the earth (which we approximate by the radius of the earth),

$$F_L = \tfrac{1}{2}C_L\,\rho A V^2, \qquad F_D = \tfrac{1}{2}C_D\,\rho A V^2,$$

C_L is the coefficient of lift, and C_D is the coefficient of drag (see Figure 2).

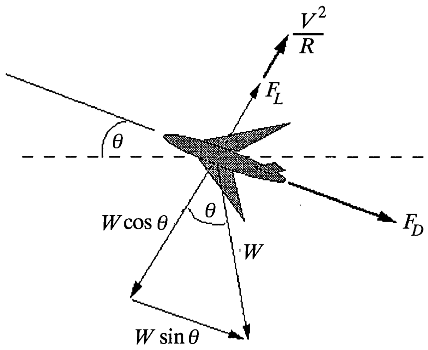


Figure 2. Forces on an unpowered aircraft

In general, even for a parachute, the equations in (10) do not have a closed form solution. However, there exists a closed form solution in one remarkable case that models re-entry of the space shuttle into the earth 's atmosphere. We discuss this example in Section 5.

5. RE-ENTRY OF THE SPACE SHUTTLE. The following model provides a reasonably accurate model for a *lifting body*, such as the space shuttle on re-entry into the atmosphere, with a closed form solution. This remarkable example should be much better known to mathematicians and can easily be presented in a first course on differential equations.

During much of the time during the space shuttle's re-entry, its velocity is approximately perpendicular to a line connecting the shuttle to the center of the earth, although at some instants the angle is quite large. In this model we assume that this is the case for all time. It then follows from (10), using $\theta = 0$, that the tangential velocity V of the shuttle satisfies

$$m\frac{dV}{dt} = -F_D, \qquad mV^2/r_E = -F_L + W, \tag{11}$$

where F_L = lift force = $C_L \rho V^2 A/2$, F_D = drag force = $C_D \rho V^2 A/2$, and r_E = radius of the earth.

For the space shuttle, it is reasonable to assume that $C_L \approx 0.5$, $C_D \approx 0.5$, $A \approx 372\,m^2$, and $W/(AC_L) \approx 100$. Over the flight envelope of the space shuttle, the quantity $F_L/F_D = C_L/C_D$ varies from about 1.0 to 1.8, and at high speeds it is roughly 1.0; for this simple example we approximate it by the constant 1.0.

We can rewrite (11) as

$$\frac{F_L}{W} = 1 - \left(\frac{V}{V_C}\right)^2 \quad \text{and} \quad \frac{F_D}{W} = -\frac{dV}{dt}\Big/ g,$$

where $V_C = \sqrt{gr_E}$. Dividing these equations gives the single equation

$$\frac{F_D}{F_L}\left(1 - \left(\frac{V}{V_C}\right)^2\right) = -\frac{dV}{dt}\Big/ g.$$

Since $F_D/F_L = C_D/C_L$, we obtain the separable equation

$$\frac{-\dfrac{dV}{V_c}}{1-\left(\dfrac{V^2}{V_c^2}\right)^2} = \frac{C_D g}{C_L V_c}\,dt,$$

which can be integrated to yield the closed form solution

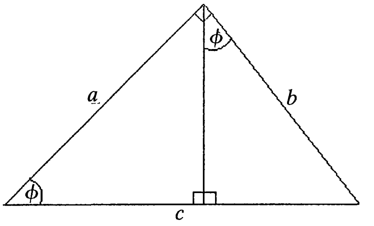
$$\begin{aligned} V(t) &= V_C \tanh\left(\frac{-C_D g}{C_L V_C}t + \operatorname{arctanh}\left(\frac{V_0}{V_C}\right)\right) \\ &= V_C \tanh\left(\frac{-C_D g}{C_L\sqrt{g r_E}}t + \operatorname{arctanh}\left(\frac{V_0}{\sqrt{g r_E}}\right)\right), \end{aligned} \tag{12}$$

where $V(0) = V_0$. Actual space shuttle flight test data **[5]** show that the velocity predicted by this simple model is reasonably accurate even though it is based on many simplifying assumptions.

One can use (12) to estimate the maximum acceleration experienced by the space shuttle upon re-entry.

6. PROOF OF THE PYTHAGOREAN THEOREM USING DIMENSIONAL ANALYSIS. We follow [**1**, p. 49] and give an insightful application of dimensional analysis to prove the Pythagorean theorem.

The area A of a right triangle is determined by its hypotenuse c and, for definiteness, the lesser of the acute angles ϕ: $A = f(c, \phi)$. Since the units of area are the square of units of length, dimensional analysis gives $A = c^2 g(\phi)$. The altitude perpendicular to the hypotenuse (see Figure 3) divides the basic triangle into two right triangles that are similar to it, and whose hypotenuses are the sides a and b of the original triangle. Their areas are $A_1 = a^2 g(\phi)$ and $A_2 = b^2 g(\phi)$. But $A = A_1 + A_2$, and thus $c^2 g(\phi) = a^2 g(\phi) + b^2 g(\phi)$. Hence $a^2 + b^2 = c^2$. ■



$a > b$

Figure 3. Right triangle

7. CONCLUSION. We have discussed models of motion for objects with *small* Reynolds numbers ($R < 1$) and for *large* Reynolds numbers ($R > \approx 100$). It is quite difficult to model the motion of most objects with Reynolds numbers in the intermediate range. The models we have discussed are quite popular with students and impress upon them, early in a differential equations course, the power of differential equations to model non-trivial physical phenomena. We applaud the trend in the new generation of calculus and differential equations texts to discuss more physical and biological models, and to make model building a major focus of the course. However, textbook writers and instructors should strive to present models based on correct physical or biological principles.

ACKNOWLEDGMENTS. The second author thanks Professor Joe Hammack for several enlightening discussions, Professor Carlos Felippa for providing a reference to the Pi theorem, Professor Howard Stone for his careful reading of a preliminary draft of this manuscript and for providing many useful comments and insights, and the following skydiving experts for their enthusiastic help: Gary Douris, George Galloway, Paul Koning, Michael Mathews, Colleen McGrath, Dan Poynter, and Jack Rumple.

REFERENCES

1. G. Barenblatt, *Scaling, Self-Similarity, and Intermediate Asymptotics*, Cambridge University Press, 1996.
2. G. Birkhoff, *Hydrodynamics*, Princeton University Press, 1960.
3. J. Boyling, A Short Proof of the Pi Theorem of Dimensional Analysis, *Z. Angew. Math. Phys.* **30** (1979) 531–533.
4. P. Brancazio, *Sport Science*, Simon and Shuster, 1984.
5. H. Compton, et al., *Shuttle (STS − 1) Entry Trajectory Reconstruction*, AIAA Paper No. 81-2459 (1981).
6. M. Denny, *Air and Water: The Biology and Physics of Life's Media*, Princeton University Press, 1993.
7. W. L. Hankey, *Re-Entry Aerodynamics*, IAA Press, 1988.
8. S. Hoerner, *Fluid-Dynamic Drag*, 2nd edition, published by author, 1958, available by calling (206) 576–3997.
9. D. Hughes-Hallet, A. Gleason, et al., *Calculus*, Wiley, 1994.
10. T. Knacke, *The Parachute Recovery Systems Design Manual*, Para Publishing, 1992.
11. L. Landau and E. Lifshitz, *Fluid Mechanics*, 2nd edition, Butterworth and Heinemann, 1987.
12. B. McCormick, *Aerodynamics, Aeronautics, and Flight Mechanics*, Wiley, 1995.
13. R. D. Mehta, *Aerodynamics of Sports Balls* Annual Review of Fluid Mechanics **17** (1985), Annual Reviews Inc.
14. W. Meyer, *Concepts of Mathematical Modeling*, McGraw-Hill, 1984.
15. D. Poynter, *The Parachute Manual*, 3rd edition, Para Publishing, 1991.
16. D. Poynter, *The Parachute Manual, Volume II*, Para Publishing, 1991.
17. I. Proudman and J. Pearson, Expansions at Small Reynolds Numbers for the Flow Past a Sphere and a Circular Cylinder, *J. Fluid Mechanics* **2** (1957) 237–263.
18. R. S. Shevell, *Fundamentals of Flight*, Prentice-Hall, 1983.
19. P. A. Thompson, *Compressible Fluid Dynamics*, McGraw-Hill, 1972.
20. S. Vogel, *Life in Moving Fluids: The Physical Biology of Flow*, second edition, Princeton University Press, 1994.
21. R. von Mises and K. O. Fredrichs, *Fluid Dynamics*, Springer-Verlag, 1971.
21. F. M. White, *Viscous Fluid Flow*, 2nd edition, McGraw-Hill, 1991.

LYLE N. LONG received his D.Sc. from George Washington University under the supervision of Feri Farassat and Michael Myers while he was at the Joint Institute for Advancement of Flight Sciences at NASA Langley Research Center. He was a Senior Research Scientist at Lockheed Aeronautical Systems Company for six years. He is Professor of Aerospace Engineering at The Pennsylvania State University. His research interests include computational aeroacoustics, computational fluid dynamics, and parallel processing.
Department of Aerospace Engineering, The Pennsylvania State University, University Park, PA 16802, lnl@psu.edu

HOWARD WEISS received his Ph.D. from the University of Maryland under the supervision of Scott Wolpert. He was a Chaim Weizmann Research Fellow at the California Institute of Technology for three years and then an NSF Postdoctoral Fellow. He is Associate Professor of Mathematics at The Pennsylvania State University. His research interests include smooth dynamical systems, geodesic flows, dimension theory, and mathematical biology.
The Pennsylvania State University, University Park, PA 16802
weiss@math.psu.edu

George Green: An Enigmatic Mathematician

D. M. Cannell

The mathematics developed by George Green has been widely applied in modern physics and quantum electrodynamics yet his life has remained something of a mystery. There is no reference to him in the main volumes of the *Dictionary of Scientific Biography*: he is belatedly and briefly included in the *Supplement* of 1976. The bicentenary celebrations of his birth in 1993 focused increasing attention on the man who gave the world Green's functions.

This MONTHLY has already published an authoritative account [3] of Green's first and seminal publication, *An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism*, which examined in some detail his use and development of mathematical sources. What is presented here is a short survey of Green's life and an examination of the circumstances and the social environment in which he lived. The great problem is the lack of material concerning Green. His output was small; ten papers including the *Essay*, written in the space of eleven years, amounting to less than 250 pages of print. There are no manuscripts, no working papers, no diaries, no memorabilia. There are a dozen letters to his patron, but no replies. There is no portrait or photograph. There was no established family house for his common-law wife, Jane Smith, and their seven children, and when the last, Clara, died in 1919, the family was thought to be extinct.

The few official facts of Green's life were established and published by H. G. Green (no relation) [5]. His paper included two valuable letters written in 1845 by Green's cousin and brother-in-law, William Tomlin, and Edward Bromhead, his patron. In the 1970s Green's letters to Bromhead came to light—the only information extant in Green's own hand, the only testimony to his mathematical thought and the only revelation of his personality. Green died at the age of forty-seven, just at the time when his work was about to be recognized: in consequence he established no reputation in his lifetime.

Green was born in 1793 and died in 1841. He was the only son of a Nottingham baker, who prospered sufficiently to extend his bakery business and build his own mill in the nearby village of Sneinton. Later he built a family house next to the mill; he acquired further property and died in 1829, sufficiently affluent to allow his son George to live on his income for the rest of his life. At fourteen George was apprenticed to his father's mill manager, William Smith, father of Jane. He found his duties as a miller "irksome" but as the only son he was obliged to work for his father. By the age of eight he had shown a passion for mathematics, such that his father sent him to an expensive town school, Robert Goodacre's Academy, but he left the following year and started work in his father's bakery.

Nothing more is known of George Green's life for certain until he reached the age of thirty and became a member of the Nottingham Subscription Library. No other serious library facilities existed in Nottingham at this period, but the Subscription Library was in effect a gentlemen's club, housed in the elegant

Georgian Bromley House in the Market Square. The members owned the property, purchased books according to their taste, and confined their numbers to prosperous and "respectable inhabitants of the town" drawn from the professional classes, the gentry, and successful men of business. George Green, a working miller, and an artisan who worked with his hands, was an unlikely member of such a gathering. But apparently he was becoming known in the town for his mathematical interests; in any case he was probably sponsored by his cousin William Tomlin, a successful man of property and an active member of the Library. Five years later Green produced his first and most important work, *An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism.* Green published the *Essay* at his own expense and opened a subscription list—a usual procedure at this period. As Bromhead wrote in 1845, "I met with a subscription list, for his first mathem1. publication, and added my name as Country Gentlemen often do by way of encouraging every attempt at provincial literature" **[4a]**. So Sir Edward Bromhead, of Thurlby Hall near Lincoln, became one of the fifty-one subscribers to the *Essay*, half of whom were members of the Library.

The *Essay*, which as far as is known was Green's own unaided work, was one of great originality, bu tit attracted little attention at the time and Green despondently resumed his milling. The death of his father, and his consequent independence, was followed by his first meeting with Bromhead in 1830, who encouraged him to resume his studies. Green wrote three papers, two on electricity and magnetism (which Bromhead sponsored for publication in the *Transactions* of the Cambridge Philosophical Society), and the third on hydrodynamics in the *Transactions* of the Royal Society of Edinburgh, of which Bromhead was a Fellow, as he was of the Royal Society of London. The period 1830 to 1833 produced Green's letters to Bromhead. The two men, despite difference in rank, were similar in age and had much in common in matters mathematical, since Bromhead had a strong interest in mathematical analysis. At Cambridge he had met his lifelong friends Charles Babbage and John Frederick Herschel, and they, with George Peacock and a few others, had formed the undergraduate Analytical Society in 1812, with the aim of publicising, in Cambridge, the continental mathematics deriving from Leibniz, in preference to the 'fluxions' of his contemporary, Isaac Newton, now deified in Cambridge. This led to the publication in 1816 of their translation of Lacroix's *Traité du calcul différentiel et du calcul intégral.*

Finally, in October 1833, Green enrolled as an undergraduate in Bromhead's college of Gonville and Caius, usually referred to as 'Caius.' In January 1837 he took his mathematical examinations and emerged Fourth Wrangler. (Wrangler was the Cambridge term for first class honours men, whose names were at that time published in order of merit, the Senior Wrangler being the first, and the last being awarded the Wooden Spoon.) In the following two and a half years, Green published his six final papers in the Cambridge *Transactions*. These, like the Edinburgh paper, dealt with wave theory based on studies in hydrodynamics, sound, and light. Six years after coming to Cambridge, Green was elected a Fellow of Caius, a position that would have allowed him to stay on in Cambridge, studying, writing, and consorting with Cambridge academics. Tragically, after holding his Fellowship for only two terms, he returned in failing health to Nottingham, "with the opinion," William Tomlin relates, "that he should never recover from his illness and which became verified in little more than a year's time on 31st May 1841" **[4b]**. Green's grave in St. Stephen's churchyard, Sneinton, is just across the road from the mill where he had worked for nearly twenty years.

Green's death in Nottingham caused little stir. His Fellowship had been noted in a single sentence in the *Nottingham Review* and on his death it contributed a modest obituary which concluded: "Had his death been prolonged, he might have stood eminently high as a mathematician." It would appear that Cambridge did not long hold Green in memory either. Four years after his death, an enquiry to Caius College was passed to Bromhead, which elicited his and Tomlin's letters of 1845. The enquiry produced more than the information contained in the letters, however, invaluable though that has proved, since it led to the retrieval of the long-neglected *Essay* of 1828.

The re-discovery of the *Essay* is one of the better-known incidents in the Green story and is recounted in the biographies of both Kelvin [**11**] and Liouville [**7**]. William Thomson, the future Lord Kelvin, took his degree in January 1845 and left for Paris to spend the summer working with Victor Regnault in his 'cabinet de physique,' and with letters of introduction to the mathematicians then in Paris: Liouville, Chasles, Sturm, and others. Thomson was already keenly interested in the problems of electricity and magnetism and he was first alerted to the existence of Green's *Essay* by a footnote in a paper on integrals by Robert Murphy. Murphy had been a Fellow of Caius during Green's residence there. He had also been the 'rapporteur,' or referee, of Green's paper on ellipsoids, sent in 1833 by Bromhead to Whewell, for publication in the Cambridge *Transactions*, and Green had sent a copy of his *Essay* with the paper. Murphy's footnote was as follows:

> The electrical action in the third section, is measured by the tension which *would* be produced by an infinitely thin rod, communicating with the electrical body, by the attraction or repulsion of the matter; it is what Mr. Green, of Nottingham, in his ingenious *Essay* on this subject, has denominated the Potential Function [**8**].

Thomson immediately visited the Cambridge booksellers. Not surprisingly, they knew nothing of an *Essay* published privately in Nottingham eighteen years previously. Just by chance however, on the eve of his departure for Paris, Thomson met his tutor, William Hopkins, who passed him two copies of the *Essay*, which Green had given him. (A later paper, inscribed "W Hopkins Esquire" in Green's hand, is now in the Green Archive, in the University of Nottingham.)

"I had only time that evening to look at some pages of it, which astonished me," wrote Lord Kelvin in 1907, a fortnight before he died at the age of eighty-three. "Next day, if I remember right, on the top of a diligence on my way to Paris, I managed to read some more of it" [**4c**].

Green's *Essay* caused a sensation among the mathematicians in Paris. In it they found Green's solutions to the problems then confronting them. The *Essay* was published by Crelle in his *Journal* in three parts in 1850, 1852, and 1854. By 1900, as noted by Grattan-Guinness, Green's functions, referred to as such by Burkhardt and Meyer [**1**] and keenly propagated by Riemann and Neumann, were well known to German mathematicians. The *Essay* was not reprinted in England until 1871, in *Mathematical Papers of the late George Green*, edited by N. M. Ferrers, promoted by Caius College and printed in London.

William Thomson did more than re-discover the *Essay*. He developed a life-long admiration for Green and did much to establish his posthumous reputation. He further developed Green's theories in his own research in electromagnetism, e.g., his method of images. His friend G. G. Stokes likewise developed Green's work on

wave theory in his own studies in hydrodynamics. Green's considerable influence on the development of nineteenth century classical physics is summed up in the words of Edmund Whittaker [**12**]:

> It is no exaggeration to describe Green as the founder of that 'Cambridge School' of natural philosophers of which Kelvin, Stokes, Rayleigh, Clerk Maxwell, Lamb, Larmor and Love were the most illustrious members in the later half of the nineteenth century.

It was through the work of Julian Schwinger and Freeman Dyson, in the mid-1940s, that Green's mathematics is now used in quantum electrodynamics. His physical concepts have found their way into various branches of modern physics and are applied in many different technologies. It is part of Green's tragedy not only that he died at the time his work was about to achieve recognition but also that he died without any awareness or indication of his future greatness. There had been but one gleam of what might have been vouchsafed him in the way of a possible European reputation in his lifetime. In the mid-1970s there came to light copies of Green's papers of 1838 and 1839, now in private hands, inscribed by Green to "Prof. Jacobi from the Author"—Green's usual inscription. Jacobi in Königsberg

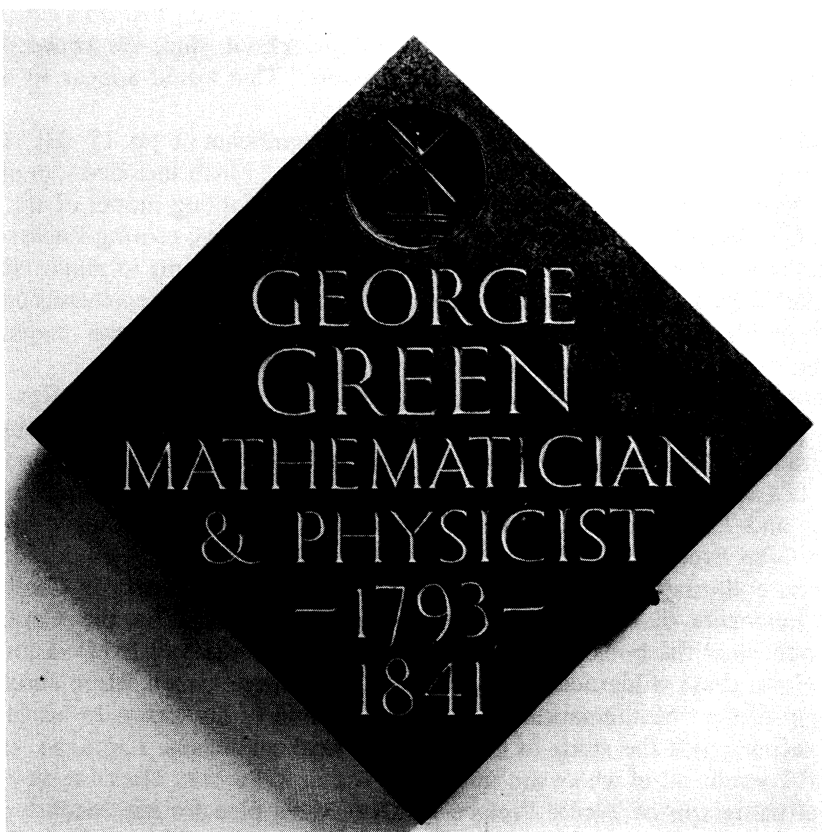


Figure 1. Plaque to George Green in Westminster Abbey, London, dedicated in July 1993 on the bicentenary of his birth.

was a regular recipient of the Cambridge *Transactions*. One wonders how Green came to send him extra copies. He would not have sent the papers unsolicited; he had neither the academic nor the social standing to do so, nor would it have been his nature.

These then are the bare bones of Green's story. Some knowledge of early nineteenth century social history and of conditions in Nottingham provide a tentative reconstruction of his environment, but how well did Green's little-known personality fit into it?

To what extent did Green's family circumstances produce or foster his genius? His grandfather was a farmer whose family had for generations worked the same acres north of Nottingham. His father, George Green, apprenticed to a baker in Nottingham, later married his employer's daughter and set up his own business nearby: nothing much here to suggest mathematical genius. On the other hand, his cousin, William Tomlin, refers to a "strong inclination for as well as a profound knowledge in the mathematics," as a result of which Green's semi-literate father in 1800 sent him to Goodacre's Academy. Young George would normally have been sent for a few terms to a 'writing school' where tradesmen's sons were taught at little expense the rudiments of reading, writing, arithmetic, and bookkeeping. In Nottingham in 1800 more general opportunities for learning were not available to the children of artisans and the poorer classes, until religious and national organisations started setting up schools in the 1830s and 40s. It says much for Green Senior that he was prepared to spend good money on his son's education, but after four terms George was withdrawn from school, since his knowledge of mathematics "far transcended that of his masters." This would appear to be the only formal education George Green ever received.

Goodacre's influence, however, may have been significant [**2**, pp. 17–23]. He was proud of his collection of "Philosophical Instruments," which included "an electrical machine," magnetic apparatus, and an orrery, or working model of the solar system. Goodacre later became a lecturer in popular science, touring England and Scotland, and from 1824 to 1827 he gave lectures on astronomy in major cities in the United States. His enthusiastic demonstrations may well have aroused Green's interest in physical phenomena, and particularly in electricity and magnetism, which he developed later in the *Essay*.

A more clearly defined influence comes from a book published by Rev. John Toplis, who in 1806 arrived from Cambridge as headmaster of the Nottingham Free Grammar School. This was a translation of the first book of the *Mécanique Céleste* of Laplace published in 1799. Toplis had been ranked Eleventh Wrangler in 1802 and had subsequently been appointed Tutor and Fellow of Queens' College. Like Bromhead and his friends in the Analytical Society a decade later, he was an enthusiastic convert from the Newtonian fluxions taught in Cambridge to the Leibnizian-based mathematical analysis widely used on the Continent. Toplis published the book at his own expense in Nottingham in 1814—a lone cry in the Nottingham wilderness, but one heard by George Green. Here surely lies the origin of the "Mathematical Analysis" in the title of his *Essay*. In his Preface Toplis recommends the study of other French mathematicians, Lagrange, Legendre, and Lacroix, all of whom are quoted by Green in his text. Green even echoes the final paragraph of Toplis' Preface in his own—a plea for the reader's indulgence for shortcomings in the work, due to the distractions of their daily occupations. Toplis' book was on sale at the Nottingham booksellers and Green obviously studied its contents. An intriguing question now arises: did the author actually

teach him? Green would have been about fourteen and starting his apprenticeship as a miller when Toplis arrived in Nottingham. He returned to Queens' in 1819 when Green was twenty-six. The Free Grammar School with its resident headmaster was less than five minutes' walk from the Green family home. Given each their passion for mathematics, quite apart from social factors such as living in the same neighbourhood, it seems quite possible that Toplis could have been Green's mentor, an interesting hypothesis for which there is only circumstantial evidence.

Apart from Goodacre and Toplis, there appears to be no further influence on Green's development until he joined the Nottingham Subscription Library in 1823. In his daily work in the mill, however, Green was well aware of the mathematics underlying its construction and mechanics. In 1828 a self-important minor mathematician from London visited Nottingham: "I heard of a young miller [Green was thirty-five!] of the name of Green, who had been printing a quarto, in which he had investigated with La Place-like precision the laws of supposed electrical action." Sir Richard Phillips was scathing on the question of contemporary physics —on "Ivory's waste of time about imaginary capillary attraction, and La Place's whimsical speculations about gravific atoms," and regrets that "Mr. Green has spent so much ingenuity in misapplying his sound mathematical learning upon it" [**10**]. Sir Richard does condescend, however, to admire Green's mathematical calculations: "His sails have a radius of twelve yards and revolve twenty-five times a minute, or more than a mile at the extremities. This great velocity carries round the stones, which are sixteen feet in circumference, 162 times a minute . . . ," and so on. Green's response to this visit can only be imagined: he may have wondered whether, in meeting Sir Edward Bromhead two years later, he would encounter a second Sir Richard Phillips.

As a member of the Bromley House Library, Green entered a world different from the one he knew. It helped both his intellectual and social development. Intellectually he was in a class of his own; in five years he would publish the *Essay*. What must be borne in mind, however, is that Green was not the usual gentleman-scholar of the period. He had neither the leisure nor the education. He was a working miller with ageing parents, a growing family, and considerable business responsibilities. It is reasonable to assume, however, that he was already familiar with the work of the French analysts advocated by Toplis. He would have absorbed much of the physics with the mathematics and appears to have fastened onto electricity and magnetism as topics for further study. So what had the Nottingham Subscription Library to offer?

Of the few titles listed under Natural Philosophy in the Library, it is unlikely that any would have been of use to Green, since at this stage he was virtually at the cutting edge of continental science in his particular field and would shortly produce his functions and his Theorem. The current taste lay in Moral Philosophy, in literature, history, theology, travel, and biography. The dozen or so books, listed under Natural Philosophy—such as Hutton's *Course of Mathematics* of 1798 and Gregory's *Treatise of Mechanics* of 1806—were written by professors at the Royal Military Academy at Woolwich and served primarily as textbooks for their students. These were doubtless helpful to some of the Library members, but it was Biot's *Traité de Physique* of 1816, relating the recent experiments in electricity of Coulomb, which Green found more useful and which he must have acquired independently. What he undoubtedly found useful were the *Transactions* of the Royal Society of London. It is obvious both from the Preface to the *Essay* and his letters to Bromhead that Green trawled through the *Transactions*. (Unfortunately

for Green, the Library did not subscribe to the *Philosophical Magazine*, which he would have found helpful, but which he had probably never heard of, the Library members preferring to read the more popular *Gentleman's Magazine* instead.)

Green's primary sources for the "Theories of Electricity and Magnetism" discussed in the *Essay* were the memoirs of Poisson. Those on electricity were published by the Institut de France in 1811 and 1812; the first of several on magnetism by the Académie des Sciences in 1821. The questions inevitably arise: how did Green in his isolation in Nottingham know of their existence and where did he find them? The answer to the first is straightforward, since they were listed in the "Presents" recorded in the annual *Transactions* of the Royal Society, to which Green had access in the Library. But where did he find them? Booksellers could order published texts, but memoirs? There were copies available to Fellows in the Library of the Royal Society in London and it was possible for Fellows to accompany, or provide a suitable letter of introduction for, a non-member. This appears to be the only way that Green would have had access to Poisson's memoirs on electricity—two memoirs "of singular elegance which to be duly appreciated [sic] must be read," states Green in the Preface to his *Essay*, and this would be no idle statement on his part. There was one Fellow of the Royal Society in Nottingham at this time. Dr. John Storer had been the first President of the Bromley House Library from 1816 to 1821 and was a well known local figure. Earlier he had been instrumental in setting up the new Lunatic Asylum in Sneinton, where old George Green was by then a prosperous miller and a churchwarden of St. Stephen's. Could Dr. Storer have provided George Green with a letter of introduction to the Librarian of the Royal Society?

Another source lies in the abstracts published in the scientific journals. The Library subscribed to three such journals prior to 1828. An entry in the minutes of 1825 records a decision to subscribe to Brewster's *Edinburgh Journal of Science*, founded the previous year, and it may be presumed that Thompson's *Annals of Philosophy* and the *Quarterly Journal of Science and the Arts* had been taken from an earlier date. The latter published in 1824 an extended extract in translation of Poisson's memoir on magnetism of February of that year and read to the Royal Academy of Science, and also a second, in 1825, of a further extract of his memoir of December 1824, though it is the earlier memoirs of 1821 and 1822 to which Green makes more frequent references in the *Essay* [**9**]. Presumably he already had access to these, as he had to the earlier memoirs on electricity. Apart from the memoirs of Poisson, it would appear that Green's knowledge of earlier writing on electricity and magnetism was otherwise confined to the few papers in the Royal Society *Transactions*. He opens the *Essay* with a reference to Cavendish's paper on electricity of 1771 (in which he identifies and corrects an unsatisfactory proposition), but as "[L]ittle appears to have been effected in the mathematical theory of electricity" since then, he moves straight to Poisson's Memoirs of "about 1812."

A more subtle influence on Green's development may be found in the community and activities of the Bromley House Library itself [**6**], [**2**, pp. 43–58]. Green's five years' membership of the Library up to 1828, and his association with its more academic and professional members, would have provided a valuable contrast to his normal working life and his contacts with tradespeople and labourers. The Library's second President, Rev. White Almond, was, like Toplis, a mathematics graduate from Queens' Cambridge; they were both local men of similar age and their lives in Nottingham overlapped from 1814 to 1819. Almond was a member of the Royal Astronomical Society; he owned a telescope and was an eager participant in the Debating Society, which also met in Bromley House. Dr. Alexander

Manson was a Fellow of the Royal Society of Edinburgh and one of the town's leading physicians; he pioneered the use of iodine in surgery. There were several clergymen, either in benefices or proprietors of schools. The Library had been formed primarily for the edification of its members, but it soon became the town's centre of culture and scientific interest. It sponsored lectures, enlivened by models and demonstrations, on chemistry, electricity and magnetism, mechanics, and astronomy, and it hosted itinerant lecturers, of whom Robert Goodacre became a successful example. In response to growing working class aspirations and demands for education, the more philanthropic members of the Library were instrumental in setting up the Mechanics Institute in 1824 and later the 'Artizans Library.' Green's membership thus coincided with a period of considerable interest in the promotion of popular scientific knowledge, which possibly accounted for the support of the twenty-five members who subscribed to the publication of his *Essay*.

Green, with a modest local reputation but no influence, contacts, or sponsor, followed the example of Goodacre (with whom it would appear he may have kept in touch), and Toplis (whom he may have got to know quite well), and published his work at his own expense. He knew well what he was doing and realised only too clearly his position. As he wrote to Bromhead in 1830:

> Indeed the trifle [the *Essay*] would never have appeared before the public as an independent work if I had then possessed the means of making its contents known in any other way but as I thought it contained something new and feared that coming from an unknown individual it might not be deemed worthy of the notice of a learned society I ventured to publish it at my own risk feeling conscious at the same time that this would be attended with certain loss **[4d]**.

The title page announces that the *Essay* was printed for the author by T. Wheelhouse, and sold by various booksellers in London, Cambridge (Deighton's), and Nottingham. By this date, such information was possibly a token insertion for local prestige. The London and Cambridge firms were the agents for supplying books ordered by customers at the Nottingham booksellers and printers (often the same people), but it does not imply automatic or reciprocal distribution of published texts. The *Essay* is not listed in Deighton's catalogues, and nearly twenty years later, as William Thomson found, they had not heard of it. As for publicity, notices in the *Nottingham Journal* and the *Nottingham Review* produced fifty-one subscribers, all but half a dozen from Nottingham.

All these considerations make a sad case for Green but, with his daily responsibilities of work and family, and with his lack of informed academic and social contacts, it is difficult to see what he could have done in his circumstances. At least he had the initiative, and fortunately the finance, to go for private publication.

It was the "Mathematical Analysis" in the title of Green's *Essay* that presumably provoked Bromhead's interest, in view of his Cambridge experiences. Green sent Bromhead his copy on April 19th 1828 with a note of thanks and appreciation. The only extant example of Bromhead's participation in the correspondence is the draft of a reply, found among his letters, offering to sponsor the publication of any further papers in one of the learned societies since, as he wrote in 1845, he realised the *Essay* "must be a complete failure and dead born." It was nearly two years before Bromhead received an answer to his offer, which, apart from the brief note sent with the *Essay*, was the first letter of the dozen Green would write in the

next four years:

Sneinton near Nottingham Jan 19th 1830

Sir

From some observations made to me last Saturday by Mr. Kidd of Lincoln I find that I have unintentionally been guilty of a gross neglect on an occasion where of all others I would most carefully have avoided it and therefore hope you will pardon the liberty I am about to take in endeavouring to explain the circumstance of my not having answered your very obliging and condescending letter and this explanation I am the more desirous to enter into because nothing connected with the publication of my little Essay has afforded me so much satisfaction as that it should have been found in any degree worthy of your notice.

Had I followed my own inclination I should immediately have written in order to have expressed in some measure my gratitude for the very handsome offer with which you had honored me but on mentioning my intentions to a gentleman on whose opinion I had at an early age been accustomed to rely he assured me that no answer would be expected but that on the contrary it would be considered as a liberty to trouble one so much my superior farther until I should be able to avail myself of your kind offer by forwarding some memoir to be communicated to one of the Royal Societies and as this gentlemen had seen more of the world than myself I yielded to his opinion though with reluctance lamenting at the same time that custom should compel me to act in a way so much at variance with my own feelings.

Although from a mistaken notion of propriety I have been so long hindered from making any acknowledgement for the very handsome offer you were so kind as to make I trust you will do me the justice to believe that I have felt most sensibly the honor conferred upon me by so much condescension on your part and that I have always esteemed that offer as most valuable **[4d]**.

Green then continues by describing his dilemma regarding publication, as quoted earlier, and finishes by promising to send a paper, a promise he confirms in his next letter of 13 February. The gentleman whose misguided advice Green followed was probably Robert Goodacre, recently returned from his American lecture tour; Mr. Kidd of Lincoln has proved unidentifiable.

Green visited Bromhead at Thurlby Hall, and his subsequent letters record the writing and publication, sponsored by Bromhead, of the three papers on topics associated with the *Essay* already mentioned. With this encouragement Green's aspirations grew. It would appear from both Tomlin's and Bromhead's letters of 1845 that Green had for some time been considering, at the suggestion of "[S]everal kind and respected friends" that "he should adopt an University education." Green first raised the issue with Bromhead in April 1833 and by the following June had made up his mind. Bromhead offered sponsorship to his own college and on October 1st Green entered Caius College, Cambridge. This was not an easy decision for Green. He could now sell the milling business, while retaining ownership of the mill, and live on his income. He could establish Jane Smith and their four children nearby and use the rents of the family house and mill as additional finance. But entry to Cambridge and the acquisition of a degree presented considerable hurdles to be overcome. Apart from social disadvantages (when four years later he graduated, as Fourth Wrangler, his success was recorded

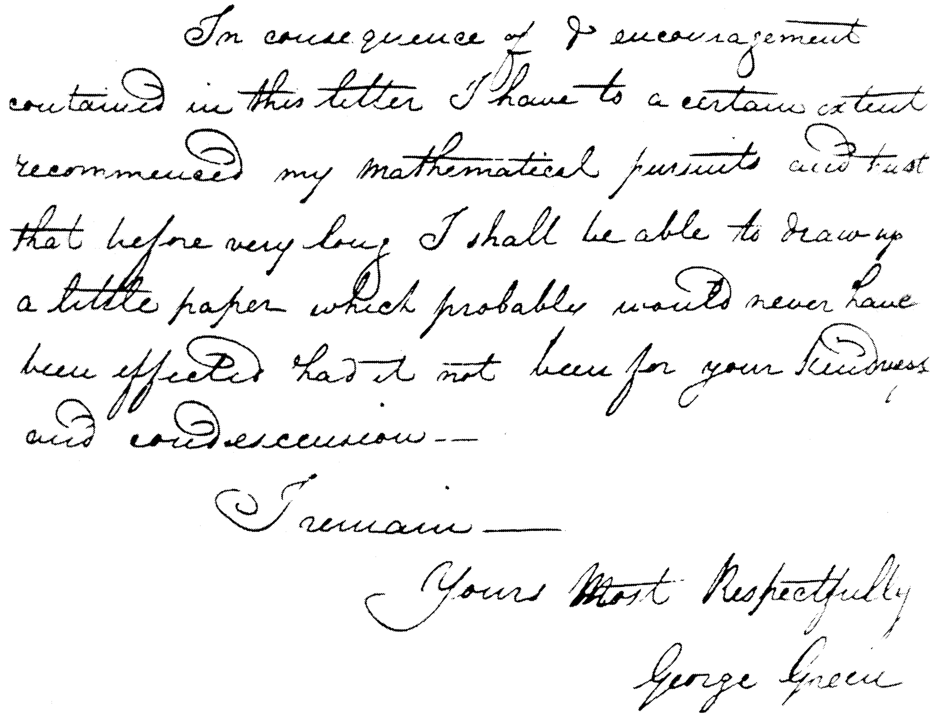

In consequence of the encouragement
contained in this letter I have to a certain extent
recommenced my mathematical pursuits and trust
that before very long I shall be able to draw up
a little paper which probably would never have
been effected had it not been for your kindness
and condescension—
I remain—
Yours Most Respectfully
George Green

Figure 2. Concluding paragraph of Green's second letter to Bromhead of 13 February 1830: *In consequence of the encouragement contained in this letter I have to a certain extent recommenced my mathematical pursuits and trust that before very long I shall be able to draw up a little paper which probably would never have been effected but for your kindness and condescension—I remain—Yours most respectfully —George Green.*

by Romilly, a celebrated Cambridge diarist as "Green of Caius (son of a miller) who was expected to be Senior Wrangler but was only fourth)," his lack of formal education was a further disadvantage. All graduates had to pass examinations in Latin and Greek before proceeding to the third year degree course, elementary enough for young twenty year olds, ex-pupils of grammar and independent schools or private home tutors, but for Green a considerable obstacle. Among some old mill accounts was found one scrap of paper in Green's writing unrelated to mill business—"*Cesar scribere et legere simul dictate et audire solebat*" (Caesar was accustomed at the same time to write, read, dictate, and listen)—poignant evidence of Green's slow climb to some level of classical literacy. His cousin William Tomlin summed up the situation when recalling Green's attendance at Goodacre's Academy:

> His schoolmasters soon perceiving this strong inclination for as well as profound knowledge in mathematics and which far transcended their own, relinquished the direction of his studies and in consequence his literary acquirements were not properly promoted: in this respect, he had when contemplating the probability of going to the University to pay some attention in his more mature years **[4e]**.

This problem still loomed large in Green's mind in 1833. He wrote to Bromhead in April: "...you are aware that I have an inclination for Cambridge if there was a

fair prospect of success. Unfortunately, I possess little Latin, less Greek, have seen too many winters, and am thus held in a state of suspense by counteracting motives" **[4f]**. The following month he asked Bromhead which college he thought suitable "for a person of my age and imperfect Classical Attainments."

Bromhead did not confine his partronage of Green merely to recommending him to go to Caius College.

> I also gave him letters of introduction to some of the most distinguished characters of the University that he might keep his object steadily in view under some awe of their names and look upwards, not of course with any view of trespassing on the social distinctions of our University, in my time much more marked than at present, but that he might venture to ask advice under any emergency **[4g]**.

Green may not have needed Bromhead's letters other than to indicate his arrival in Cambridge. This again was a custom of the time. William Thomson in 1845 arrived in Paris with a dozen or so. In fact Green came already with something of a reputation. He had had one paper published in the Cambridge *Transactions* of 1833 "Communicated by Sir Edward French Bromhead, Bart. [Baronet], M.A., F.R.S.L. and E." and a second under his own name "George Green Esq., Caius College." This latter was against all precedent since the Cambridge Philosophical Society published papers only by graduates, unlike the *Cambridge Mathematical Journal*, which included papers by undergraduates, such as William Thomson, and outsiders, such as George Boole.

Green soon raised the expectation of being the Senior Wrangler of his year (as would William Thomson, though he came second in 1845), and he is noted by Harvey Goodwin in the first edition of the *Dictionary of National Biography* as "standing head and shoulders above all in and outside of the University." In January 1837, Green was Fourth, being beaten by three young men from St. John's College, including James Joseph Sylvester, pupils of the renowned coach John Hymers. William Hopkins, already mentioned as William Thomson's tutor, was another highly respected Cambridge coach who later became Professor of Geology. The Cambridge coaches took their students through all possible examination questions. Since the Wrangler's rank depended on the maximum number of questions answered and thus marks gained, without the necessity of giving much mathematical thought to the problem set, it is perhaps not surprising that in the event George Green, with his own brand of mathematical genius, did not gain first place.

Green stayed on in Cambridge awaiting election to a Fellowship. As a graduate he could now become a Fellow of the Cambridge Philosophical Society. During the two and a half years' interval he wrote and published his six further papers on wave theory in the *Transactions*—on one occasion at least reading a paper at the same meeting as William Whewell. He was now working in the main stream of Cambridge—and indeed European—research, as the papers sent to Jacobi and references to a younger generation of French scientists, such as Fresnel and Cauchy, would suggest. In the light of his current research, it is perhaps not surprising that Green appears to have given little thought to the *Essay* of nearly twenty years earlier, despite his quiet distribution of copies to Hopkins and to the Cambridge Philosophical Society and Caius College Libraries. As it was, Thomson's youthful enthusiasm ensured re-publication of the *Essay* on the Continent, but he confessed to a lifelong regret that as editor he had not thought to reprint the *Essay* in

the *Cambridge and Dublin Mathematical Journal*, successor to the *Cambridge Mathematical Journal* **[4h]**.

Green's last recorded letter to Bromhead is dated May 22nd 1834, written in the middle of his first year examinations. After discussion of the publication of his second paper, published the following year, he concludes:

> I am very happy here and I fear too much pleased with Cambridge. This takes me in some measure from those pursuits which ought to be my proper business but I hope on my return to lay aside my freshnesses and become a regular steady Second Year man **[4i]**.

This appears to be the last written communication between the two men, though Green may have visited Bromhead at Thurlby in vacations. The latter may have thought his obligations as patron were discharged. Certainly in his letter of 1845 he makes no reference to events in Cambridge other than mentioning the letters of introduction and concludes abruptly and rather disconcertingly:

> So much for my knowledge of poor Green, but I have written to a gentleman in Nottingham, who may perhaps supply further particulars.... .

This led at one remove to William Tomlin's letter of the same date. But a closer look at the custom of patronage in this period serves to put Green's situation in longer focus and at the same time allows readers to adjust perceptions of the personalities of the two men involved. Patronage was a long-established custom. The more enlightened patron took it as a responsibility of his social position, as Bromhead certainly did, in helping George Green, and later, the young Lincoln schoolmaster, George Boole. When Bromhead received his copy of the *Essay*, he had difficulty in phrasing his offer to sponsor the publication of any later papers, since he did not know the social status of the author. Realising that the Director of the Lincoln Lunatic Asylum, of which he was Vice-President, had Nottingham connections, he wrote to him for information. The answer re-assured Bromhead as to the propriety of his actions; at the same time it offers posterity a further account of Green's reputation in Nottingham at the time of the publication of the *Essay*:

> 1828
>
> Sir
> I learn from Nottingham that Mr G Green is the Son of a Miller, who has had only a common education in the Town, but has been ever since his mind could appreciate the value of learning immoderately fond of Mathematical pursuits, and which attainments have been acquired wholly by his own perseverance unassisted by any Tutor or Preceptor: he is now only 26 or 27 years of age of rather reserved habits attends the business of the Mill, but yet finds time for his favorite Mathematical reading—
>
> Your obt. Servant
> Thos Fisher
>
> Asylum May 10 **[4j]**

Conventional epistolary forms duly reflected the differences in social status. Thus Bromhead's apparent arrogance can be viewed as a natural assumption of superiority—and responsibility—and Green's apparent obsequiousness as a normal expression of respect to one "so much [his] superior." Bromhead found it easy to offer

help, but Green had difficulty in accepting it, as shown by his reply to an unexpected invitation from Bromhead:

> You were kind enough to mention a journey to Cambridge on June 24th to see your friends Herschell Babbage and others who constitute the Chivalry of British Science. Being as yet only a beginner I think I have no right to go there and must defer that pleasure until I shall have become tolerably respectable as a man of Science should that day ever arrive.
>
> I remain with the Greatest Respect
> Yours Sincerely
> Geo: Green **[4k]**

A modern reader might assume that the association between the two mathematicians had established a certain degree of equality. Not so, however. Bromhead's letter of 1845, mentioning the letters of introduction, and in particular, the final reference to "poor Green"—with the significant omission of title—reveals clearly enough the true situation, one reflected in the meeting of Green and Sir Richard Phillips, though with the overtones of ignorance and complacency absent in the case of Bromhead.

A more attractive feature of patronage was the factor of disinterestedness, since each side was activated by a common interest in science and each served its advancement. There are two interesting reflections of this approach in the Bromhead correspondence. In his third letter to Bromhead, written in May 1832, Green writes:

> I cannot conclude without expressing my gratitude for your kind assistance which can only have arisen from a liberal desire to forward the interests of science by encouraging even the most humble cultivators
>
> I remain with the greatest respect
> Yours Very Sincerely
> Geo: Green **[4l]**

Bromhead expresses the same sentiment in his letter to Whewell of November of that year when discussing editorial changes to Green's paper on the equilibrium of fluids:

> Would it be too great a favor to request that you would become Gardener in this pruning.... Mr. Green had retired in despair from mathematics and undertook this memoir at my request, from which you will see that a little encouragement may secure him as a Recruit to the very small troop who serve under the severe sciences... **[4m]**.

The story of Bromhead's patronage highlights Green's dilemma. Given his social status, his personal circumstances, his self-confessed limited knowledge of the world, and his lack of academic contacts, it seems evident that Green could never have made his way alone. Bromhead's patronage was vital to his work and intellectual advancement. Unfortunately both were cut short by Green's early death in 1841.

Green's posthumous reputation proved as varied and uncertain as his life. Reference has already been made to his lack of reputation in Nottingham and to his irregular family situation. These are closely linked. Green never openly

acknowledged the existence of Jane Smith and their seven children, though they were well provided for in his will, and, until his final return from Cambridge, Green and Jane Smith are not known to have lived under the same roof. Green died, however, in the house where Jane then lived at 3 Notintone Place, opposite St. Stephen's Church and a stone's throw from the mill. Mrs. Jane Green, as she had always been known, continued to live in the house after Green's death and was later buried in a grave adjoining his. Local hearsay reported that Green Senior had opposed marriage; later, Green's aspirations for Cambridge and a Fellowship dictated celibacy, since prospective Fellows had to be unmarried. Furthermore, Victorian society imposed a strict moral code under which illegitimacy was a social disgrace and Green's surviving children, Jane, George, Elizabeth, and Clara, while

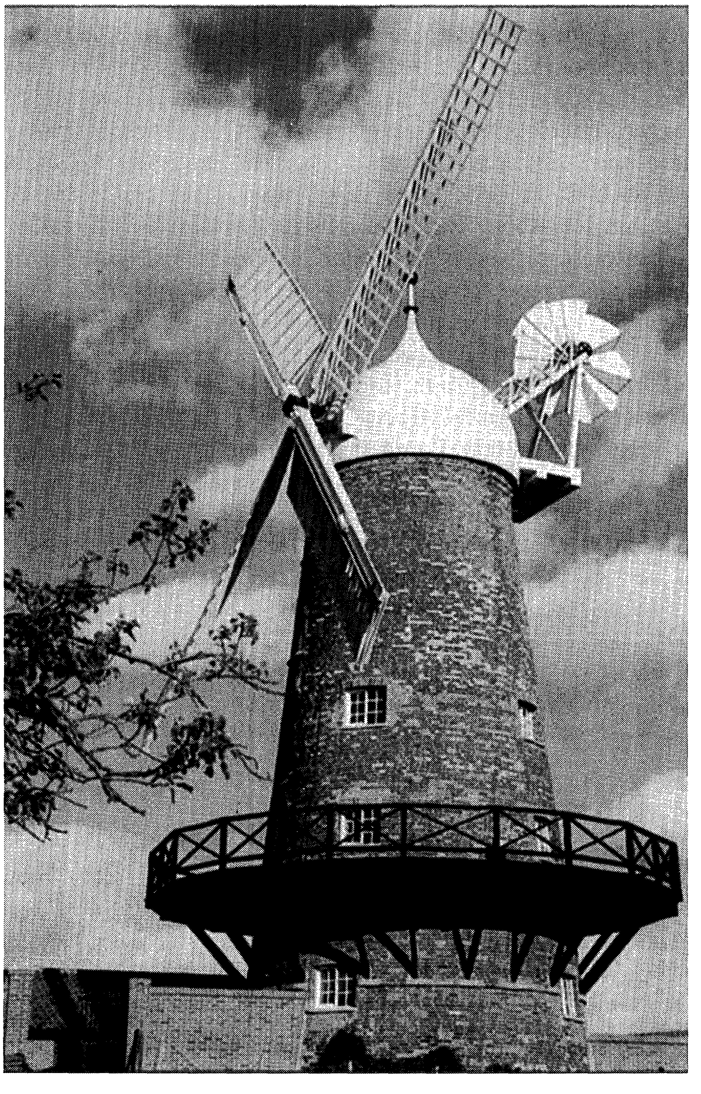


Figure 3. Green's Mill in Sneinton, Nottingham, built in 1807 by his father, where George Green laboured for over twenty years. It was restored in 1985 as a memorial to him by the George Green Memorial Fund under its founder-chairman Professor Lawrie Challis.

living comfortably, bore its stigma. Jane and Elizabeth married, but George, who took a mathematics degree at St. John's, Cambridge in his late twenties, was obliged to hide his identity so as not to compromise the memory of his father, and committed suicide in London at the age of forty—a further social disgrace at this period.

Only the youngest, Clara, remained, dying in the poor hospital in 1919 aged seventy-eight. She lies buried in a corner of the town cemetery and her forlorn and neglected grave was discovered only a few years ago. She inherited finally all of Green's once-prosperous estate of 1841, the mill, the family house, and land, together with property in Sneinton and Nottingham, but all had been heavily mortgaged and she left only debts. On her death the family was presumed extinct, and the Crown disposed of her effects. The mill was in a ruined state, not having been worked since the 1860s. It was restored to full working order in 1985 by the George Green Memorial Fund. Clara Green, a colourful and eccentric figure, reputedly lived in a hut in the garden of one of the Sneinton houses she owned and on her death neighbours cleared out all the rubbish and papers and burnt them. This could possibly account for the fact that none of Green's writings has survived.

One can now appreciate why Green's name and reputation were unrecognised in Nottingham until the restoration of his mill and the celebrations of the bicentenary of his birth in 1993. These culminated in the dedication of a plaque to Green in Westminster Abbey in London. This lies next to Newton's grave in the Sanctuary and in close proximity to similar plaques to Faraday, Kelvin, and Clerk Maxwell. The celebrations were attended by fourteen blood descendants of George Green, two from Canada and one from New York, with members of their families, twenty five in all. This surprising twist to the Green story is explained by the fact that Green's eldest daughter Jane, the only one of the seven children to have offspring, left one son, George Green Moth. His six children were born in two marriages, at forty years' interval, and the two families, who grew up in ignorance of each other's existence, were traced in the 1970s and 80s. They were united for the first time in Nottingham at the Civic Service of Thanksgiving in St. Stephen's Church on 13 July 1993, the eve of the bicentenary of their ancestor's birth.

REFERENCES

1. Burkhardt, H. and Mayer, F. W. F. 1900. Potentialtheorie, in *Encyklopaedie der mathematischen Wissenschaften*, 1900, vol. 2, pt. A. 464-503 (article IIA7b).
2. Cannell, D. M., *George Green, Mathematician and Physicist, 1793–1841*, Athlone Press, London, 1993.
3. Grattan-Guinness, I., Why did George Green write his *Essay* of 1828 on Electricity and Magnetism? *Amer. Math Monthly* 102 (1995) 387–396.
4. The following extracts are from letters in *George Green: A Catalogue of Books and Manuscripts associated with Green*, Department of Manuscripts & Special Collections, University of Nottingham. For longer extracts and discussion of contents and contexts, see [2].
   a. G.G. Cat. 4A 5: Bromhead's letter of April 13th 1845 to Caius College Cambridge.
   b G.G. Cat. 4A 6ii: William Tomlin (per J. C. Williams) to Bromhead April 10th 1845.
   c G.G. Cat. 4A: Kelvin to Sir Joseph Larmor, November 19th 1907.
   d G.G. Cat. 4B2: Green to Bromhead, January 19th 1830.
   e G.G. Cat. 4A 6ii: Tomlin 1845.
   f G.G. Cat. 4B 16: Green to Bromhead, April 13th 1833.
   g G.G. Cat. 4A 5: Bromhead 1845.
   h G.G. Cat. 4A: Kelvin 1907.
   i G.G. Cat. 4B 22: Green to Bromhead, May 22nd 1834.
   j G.G. Cat. 4B 24: Thomas Fisher to Bromhead, May 10th 1828.
   k G.G. Cat. 4B 22: Green to Bromhead, May 1833.

l G.G. Cat. 4B 4: Green to Bromhead, May 17th 1832.
m G.G. Cat. 4B 8: Bromhead to William Whewell, November 26th 1832.

5. Green, H. G., Biography of George Green, Mathematical Physicist of Nottingham and Cambridge, 1793–1841. In *Essays and Studies in the History of Science and Learning* ed. M. A. Montague, Schuman, New York, 1945, pp. 544–592.
6. Inkster, I., Scientific Culture and Education in Nottingham 1800–1843. *Thoroton Society Transactions* 1978 (University of Nottingham Hallward Library, Local Studies Department): [**2**, pp. 43–58].
7. Lützen, J., *Joseph Liouville*, Springer, New York, 1990, pp. 135–141.
8. Murphy, R., On the inverse method of definite integrals with physical applications, *Cambridge Philosophical Transactions*, IV (1833) 353–408.
9. *Quarterly Journal of Science and the Arts*, 17 no. 34 (1824) 317–34, Extract of a memoir on the Theory of Magnetism by M. Poisson, read to the Royal Academy of Science 2nd February 1824, 19 no. 39 (1825) 122–32, Second memoir on the Theory of Magnetism by M. Poisson, read to the Royal Academy of Science, 27th December 1824.
10. Syer, G., A Visit to Nottingham in 1828, in *Nottinghamshire Historian* 1994. Nottingham County Library, Local Studies Department.
11. Thompson, S. P., *Life of Lord Kelvin*, London, Macmillan, 1910, Vol. I, pp. 115–119.
12. Whittaker, E., *History of the Theories of Aether and Electricity*, London, Nelson, 1910, p. 153.

MARY CANNEL, born in Liverpool, U.K., confesses to having had the temerity to write the biography of George Green, although she is neither a mathematician nor a physicist. She graduated from Liverpool University in French and History, and for some years taught students in schools and colleges in England and abroad, before embarking on the training of teachers. She retired as Principal of the College of Education in Nottingham, where she became interested in the life of the town's long neglected genius. She published her biography of Green in 1993, the year of the bicentenary celebrations of his birth. She is the Secretary of the George Green Memorial Fund, based in the Physics Department of Nottingham University. The Fund instigated the restoration of Green's Mill in Nottingham and the dedication of a plaque in his honour in Westminster Abbey in London. For more about George Green, mathematician and physicist, visit http://www.nottingham.ac.uk/nppzwww/green/.

c/o Mrs. Linda Wightman, Physics Department, University of Nottingham, Nottingham, NG7 2RD, U.K.
ppzlaw@ppn1.nottingham.ac.uk

NOTES

Edited by **Jimmie D. Lawson and William Adkins**

Magic "Squares" Indeed!

Arthur T. Benjamin and Kan Yasuda

1 INTRODUCTION. Behold the remarkable property of the magic square:

$$\begin{bmatrix} 6 & 1 & 8 \\ 7 & 5 & 3 \\ 2 & 9 & 4 \end{bmatrix}$$

$$\begin{aligned} 618^2 + 753^2 + 294^2 &= 816^2 + 357^2 + 492^2 \quad \text{(rows)} \\ 672^2 + 159^2 + 834^2 &= 276^2 + 951^2 + 438^2 \quad \text{(columns)} \\ 654^2 + 132^2 + 879^2 &= 456^2 + 231^2 + 978^2 \quad \text{(diagonals)} \\ 639^2 + 174^2 + 852^2 &= 936^2 + 471^2 + 258^2 \quad \text{(counter-diagonals)} \\ 654^2 + 798^2 + 213^2 &= 456^2 + 897^2 + 312^2 \quad \text{(diagonals)} \\ 693^2 + 714^2 + 258^2 &= 396^2 + 417^2 + 852^2 \quad \text{(counter-diagonals)}. \end{aligned}$$

This property was discovered by Dr. Irving Joshua Matrix **[3]**, first published in **[5]** and more recently in **[1]**. We prove that this property holds for *every* 3-by-3 magic square, where the rows, columns, diagonals, and counter-diagonals can be read as 3-digit numbers in *any* base. We also describe n-by-n matrices that satisfy this condition, among them all circulant matrices and all symmetrical magic squares. For example, the 5-by-5 magic square in (1) also satisfies the square-palindromic property for every base.

$$\begin{bmatrix} 17 & 24 & 1 & 8 & 15 \\ 23 & 5 & 7 & 14 & 16 \\ 4 & 6 & 13 & 20 & 22 \\ 10 & 12 & 19 & 21 & 3 \\ 11 & 18 & 25 & 2 & 9 \end{bmatrix} \tag{1}$$

We must be careful when we read these numbers. The base 10 number represented by the first row of (1) is $17 \cdot 10^4 + 24 \cdot 10^3 + 1 \cdot 10^2 + 8 \cdot 10 + 15 = 194195$. The base 10 number based on the first row's reversal is 158357.

2 SUFFICIENT CONDITIONS. We say that a real matrix is *square-palindromic* if, for every base b, the sum of the squares of its rows, columns, and four sets of diagonals (as in the previous examples) are unchanged when the numbers are read "backwards" in base b. We can express this condition using matrix notation. Let M be an n-by-n matrix. Then the n numbers (in base b) represented by the rows of M are the entries of the vector $M\mathbf{b}$, where $\mathbf{b} = (b^{n-1}, b^{n-2}, \dots, b, 1)^T$, and T denotes the transpose operation. The sum of the squares of these numbers is

$$(M\mathbf{b})^T(M\mathbf{b}) = \mathbf{b}^T(M^TM)\mathbf{b}.$$

Next, the n numbers represented by the rows when read "backwards" are the entries of $MR\mathbf{b}$ where the n-by-n *reversal matrix* $R = [r_{ij}]$ has $r_{ij} = 1$ if $i + j = n + 1$, and $r_{ij} = 0$ otherwise. Note that $R^T = R^{-1} = R$. The sum of the squares of these numbers is

$$(MR\mathbf{b})^T(MR\mathbf{b}) = \mathbf{b}^T(R(M^TM)R)\mathbf{b}.$$

Hence a sufficient condition for the *rows* of M to satisfy the square-palindromic property is simply $R(M^TM)R = M^TM$. Matrices A that satisfy $RAR = A$ are called *centro-symmetric* [**6**]: $a_{ij} = a_{n+1-i,n+1-j}$. Matrices A that satisfy $RAR = A^T$ are called *persymmetric* [**4**]: $a_{ij} = a_{n+1-j,\, n+1-i}$. It is easy to see that symmetric matrices that are centro-symmetric must also be persymmetric. Since M^TM is necessarily symmetric, our sufficient condition says that M^TM is centro-symmetric, or equivalently, that

$$M^TM \text{ is persymmetric.}$$

The square-palindromic condition for the *columns* of M is the square-palindromic condition for the rows of M^T. Hence it suffices to require that

$$MM^T \text{ is persymmetric.}$$

For the first set of *diagonals*, we create a matrix $\tilde{M}$ with the property that each column of $\tilde{M}$ represents a diagonal starting from the first row of M. To do this, we introduce two other special square matrices. Let $P_k = [p_{ij}]$ denote the n-by-n *projection matrix* whose only non-zero entry is $p_{kk} = 1$. Notice that $P^T = P$, and P_kM preserves the kth row of M but turns all other rows to zeros. Let $S = [s_{ij}]$ denote the n-by-n *shift operator* where $s_{ij} = 1$ if $i - j \equiv 1 \pmod{n}$, $s_{ij} = 0$ otherwise.

The following properties of S are easily verified: $S^n = I_n$, $S^{-1} = S^T = RSR$, and MS^k shifts the columns of M over "k steps to the left". Now define

$$\tilde{M} = \sum_{i=1}^{n} P_iMS^{i-1}.$$

Hence the i-th diagonal of M, starting from the first row becomes the i-th column of $\tilde{M}$. By the column condition, these diagonals satisfy the square-palindromic property if the (i,j) entry of $\tilde{M}\tilde{M}^T$ equals its $(n+1-j, n+1-i)$ entry.

We have

$$\tilde{M}\tilde{M}^T = \sum_{i=1}^{n} P_iMS^{i-1}\left(\sum_{j=1}^{n} P_jMS^{j-1}\right)^T = \sum_{i=1}^{n}\sum_{j=1}^{n} P_iMS^{i-j}M^TP_j.$$

It follows that $\tilde{M}\tilde{M}^T$ has the same (i,j) entry as $MS^{i-j}M^T$, and the same $(n+1-j, n+1-i)$ entry as well; if $MS^{i-j}M^T$ is persymmetric, then these entries are equal. Consequently, these diagonals obey the square-palindromic property if

$$MS^kM^T \text{ is persymmetric for } k = 1, \ldots, n. \tag{2}$$

Conveniently, (2) also ensures that the counter-diagonals starting from the first row satisfy the square-palindromic property. This can be seen by mimicking the preceding explanation with $\tilde{M} = \sum_{i=1}^{n} P_iMS^{-(i-1)}$, whereby $\tilde{M}\tilde{M}^T$ has the same (i,j) and $(n+1-j, n+1-i)$ entry as $MS^{j-i}M^T$. For the other diagonal and

counterdiagonal, we obtain similar results [7], which we summarize in the following theorem:

Theorem 1. *A square matrix M has the square-palindromic property if the following matrices are all persymmetric*:

1. M^TM,
2. MM^T,
3. MS^kM^T, *for* $k = 1,\ldots,n$, *and*
4. M^TS^kM, *for* $k = 1,\ldots,n$.

3. SQUARE-PALINDROMIC MATRICES. Next we explore classes of matrices that are square-palindromic. We say that a square matrix A is *centro-skew-symmetric* if $RAR = -A$, that is, $a_{ij} + a_{n+1-i,n+1-j} = 0$.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 8 & 7 & 6 & 5 \\ 4 & 3 & 2 & 1 \end{bmatrix} \qquad \begin{bmatrix} a & b & c \\ d & 0 & -d \\ -c & -b & -a \end{bmatrix}$$

Centro-Symmetric Centro-Skew-Symmetric

Theorem 2. *Every centro-symmetric or centro-skew-symmetric matrix is square-palindromic.*

Proof: If M is centro-symmetric or centro-skew-symmetric, then the relations $RM = \pm MR$ and $R(S^k)R = S^{-k}$ ensure that M satisfies the conditions of Theorem 1. ∎

The theorem is not at all surprising since the collection of rows, columns and diagonals of M read the same backwards and forwards. The next class of matrices, however, satisfies the conditions in a non-obvious way.

We say that A is *circulant* if every entry of each "diagonal" is the same, i.e., $a_{ij} = a_{k\ell}$ if $i - j \equiv k - \ell \bmod n$ or simply $SAS^{-1} = A$. We say that A is (-1)-*circulant* if $SAS = A$.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \\ 3 & 4 & 1 & 2 \\ 2 & 3 & 4 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 1 \\ 3 & 4 & 5 & 1 & 2 \\ 4 & 5 & 1 & 2 & 3 \\ 5 & 1 & 2 & 3 & 4 \end{bmatrix}$$

Circulant (-1)-Circulant

Notice that the circulant and (-1)-circulant property is preserved under transposing. It is easy to show that the product of two circulant matrices or two (-1)-circulant matrices is circulant, while the product of a circulant and (-1)-circulant matrix is (-1)-circulant. Note that S is circulant, R is (-1)-circulant, and that all circulant matrices are persymmetric since a_{ij} and $a_{n+1-j,\, n+1-i}$ lie on the same diagonal. Consequently, if M is circulant or (-1)-circulant, the matrices M^TM, MM^T, MS^kM^T, and M^TS^kM are all circulant, and thus persymmetric. From Theorem 1, it follows that

Theorem 3. *Every circulant or (-1)-circulant matrix is square-palindromic.*

Notice that four of the six square-palindromic identities are not obvious, but two of the diagonal sums are completely trivial!

4. MAGIC AND SEMIMAGIC SQUARES. A *semi-magic square* with magic constant c is a square matrix A in which every row and column adds to c. Using matrix notation, this says that $AJ = cJ = JA$, where J is the matrix of all ones. If the main diagonal and main counter-diagonal also add to c, then the matrix is called a *magic square*. Circulant and (-1)-circulant matrices are always semi-magic, but are not necessarily magic.

A magic square A is *symmetrical* [2] if the sum of each pair of two entries that are opposite with respect to the center is $2c/n$, that is $a_{ij} + a_{n+1-i,n+1-j} = 2c/n$. Notice that a semimagic square with this property is magic.

Like the example below, magic and semi-magic squares do not necessarily satisfy the square-palindromic property.

$$\begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Semi-Magic but not square-palindromic

However,

Theorem 4. *Every symmetrical magic square is square-palindromic.*

Proof: The trick is to notice that if M is a symmetrical magic square with magic constant c, then $M = M_0 + cJ/n$, where M_0 is a symmetrical magic square with magic constant 0. But this implies that M_0 is centro-skew-symmetric. Therefore M_0 is square-palindromic and satisfies the conditions of Theorem 1. Thus, since $M_0^T M_0$ and J are persymmetric, it follows that $M^T M = (M_0 + cJ/n)^T(M_0 + cJ/n) = M_0^T M_0 + c^2 J/n$ is also persymmetric. Hence M satisfies condition 1 of Theorem 1. To verify condition 3 (the other cases are similar), notice that

$$MS^k M^T = \left(M_0 + \frac{c}{n}J\right) S^k \left(M_0 + \frac{c}{n}J\right)^T = M_0 S^k M_0^T + \frac{c^2}{n}J$$

is persymmetric for $k = 1, \ldots, n$, since M_0 satisfies condition 3 of Theorem 1. ∎

Although not all magic squares are square-palindromic, it is easy to see that all 3-by-3 magic squares are symmetrical. Consequently, we have

Theorem 5. *All* 3-*by*-3 *magic squares are square-palindromic.*

REFERENCES

1. E. J. Barbeau, *Power Play*, Mathematical Association of America, Spectrum, Washington DC, 1997.
2. W. H. Benson and O. Jacoby, *New Recreations with Magic Squares*, New York: Dover Publications, 1976.
3. M. Gardner, *Penrose Tiles To Trapdoor Ciphers ... And The Return Of Dr. Matrix*, W. H. Freeman and Company, 1989.
4. G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 1983.

5. R. Holmes, The Magic Magic Square, *Math. Gazette* LIV, **390** (1970) 376.
6. J. R. Weaver, Centro-Symmetric (Cross-Symmetric) Matrices, Their Basic Properties, Eigenvalues, and Eigenvectors, *Amer. Math. Monthly* **92** (1985) 711–717.
7. K. Yasuda, A Square Sum Property of Magic Squares, *Senior Thesis*, Mathematics Department, Harvey Mudd College, 1997.

Harvey Mudd College, Claremont, CA 91711,
benjamin@hmc.edu,
kan@msf.biglobe.ne.jp

An Elementary Proof of Binet's Formula for the Gamma Function

Zoltán Sasvári

The present note presents an elementary proof of the following important result of J. P. M. Binet [**3**, p. 249].

Theorem 1. *For* $x > 0$ *we have*

$$\Gamma(x+1) = \left(\frac{x}{\mathrm{e}}\right)^x \sqrt{2\pi x} \cdot \mathrm{e}^{\theta(x)} \tag{1}$$

where

$$\theta(x) = \int_0^\infty \left(\frac{1}{\mathrm{e}^t - 1} - \frac{1}{t} + \frac{1}{2}\right) \mathrm{e}^{-xt} \frac{1}{t}\, dt.$$

Here Γ *denotes the gamma function defined by*

$$\Gamma(x) = \int_0^\infty t^{x-1} \mathrm{e}^{-t}\, dt.$$

Since $\lim_{x\to\infty} \theta(x) = 0$, *from* (1) *we immediately obtain Stirling's formula*

$$n! = \Gamma(n+1) \sim \left(\frac{n}{\mathrm{e}}\right)^n \sqrt{2\pi n}\,.$$

Binet's formula can also be used to prove a more precise version of Stirling's asymptotic expansion

$$\log \frac{n!}{(n/\mathrm{e})^n \sqrt{2\pi n}} = \sum_{j=1}^{\infty} \frac{B_{2j}}{2j(2j-1)n^{2j-1}} = \frac{1}{12n} - \frac{1}{360n^3} + \frac{1}{1260n^5} - \cdots,$$

where the B_{2j}*'s denote the Bernoulli numbers defined by*

$$\frac{1}{\mathrm{e}^t - 1} - \frac{1}{t} + \frac{1}{2} = \sum_{j=1}^{\infty} \frac{B_{2j}}{(2j)!} t^{2j-1}.$$

For, by problem 154 *in Part I, Chapter* 4 *of* [**2**], *the inequalities*

$$\sum_{j=1}^{2N} \frac{B_{2j}}{(2j)!} t^{2j-1} < \frac{1}{\mathrm{e}^t - 1} - \frac{1}{t} + \frac{1}{2} < \sum_{j=1}^{2N+1} \frac{B_{2j}}{(2j)!} t^{2j-1}$$

hold for each nonnegative integer N. From $j! = \int_0^\infty t^j e^{-t}\,dt$ and the definition of θ we immediately obtain

$$\sum_{j=1}^{2N} \frac{B_{2j}}{2j(2j-1)n^{2j-1}} < \log\frac{n!}{(n/e)^n\sqrt{2\pi n}} < \sum_{j=1}^{2N+1} \frac{B_{2j}}{2j(2j-1)n^{2j-1}}.$$

In this MONTHLY *several derivations of Stirling's formula and asymptotic expansion have been published. We mention here only the most recent* [**1**].

To prove Binet's formula, we define the function φ by the equation

$$e^{\varphi(x)} = \frac{1}{\sqrt{2\pi}} \int_0^\infty \sqrt{x}\,(te^{1-t})^x\,dt$$

so that

$$\Gamma(x+1) = \left(\frac{x}{e}\right)^x \sqrt{2\pi x}\cdot e^{\varphi(x)}. \tag{2}$$

Binet's formula is equivalent to $\theta(x) = \varphi(x)$. We prove this equality by showing that θ and φ both satisfy a certain difference equation and that $\theta(\frac{1}{2}) = \varphi(\frac{1}{2})$.

Our first lemma tells nothing new; we present a proof for the sake of completeness.

Lemma 1. *For all $x > 0$ and $a > -x$ we have*

$$\int_0^\infty \frac{e^{-xt} - e^{-(x+a)t}}{t}\,dt = \log\left(1 + \frac{a}{x}\right). \tag{3}$$

Proof: Denoting by $f(x)$ and $g(x)$ the left and right hand sides of (3), respectively, we have $\lim_{x\to\infty} f(x) = \lim_{x\to\infty} g(x) = 0$ and $f'(x) = g'(x)$. Consequently, $f(x) = g(x)$ for all $x > 0$. ∎

Lemma 2. *For all $x > 0$ we have*

$$\varphi(x) - \varphi(x+1) = \theta(x) - \theta(x+1) = \left(x + \frac{1}{2}\right)\log\left(1 + \frac{1}{x}\right) - 1. \tag{4}$$

Proof: Denote by $g(x)$ the right-hand side of (4). That $\varphi(x) - \varphi(x+1) = g(x)$ follows immediately from (2) by using the equation $\Gamma(x+2) = (x+1)\Gamma(x+1)$. To prove the statement about θ, first note that $\lim_{x\to\infty} \theta(x) - \theta(x+1) = \lim_{x\to\infty} g(x) = 0$. Moreover,

$$\theta'(x) - \theta'(x+1) = \int_0^\infty \frac{e^{-xt} - e^{-(x+1)t}}{t} - \frac{e^{-xt} + e^{-(x+1)t}}{2}\,dt.$$

Applying (3), we obtain

$$\theta'(x) - \theta'(x+1) = \log\left(1 + \frac{1}{x}\right) - \frac{1}{2}\left(\frac{1}{x} + \frac{1}{x+1}\right) = g'(x).$$

Since the limits at ∞ and the derivatives are equal, $g(x) = \theta(x) - \theta(x+1)$. ∎

Remark. Differentiating under the integral sign in the previous proofs can be avoided by replacing $\frac{1}{t}$ by $\int_0^\infty e^{-st}\,ds$ and then using Fubini's theorem.

Lemma 3. $\varphi(\frac{1}{2}) = \theta(\frac{1}{2}) = \frac{1}{2} - \frac{1}{2}\log 2$.

Proof: Since $\Gamma(\frac{3}{2}) = \frac{1}{2}\sqrt{\pi}$, (2) yields $\varphi(\frac{1}{2}) = \frac{1}{2} - \frac{1}{2}\log 2$.

As to $\theta(\frac{1}{2})$, we follow an idea of A. Pringsheim [3, p. 249]. By an obvious substitution,

$$\theta(1) = \int_0^\infty \left(\frac{1}{e^{\frac{1}{2}t} - 1} - \frac{2}{t} + \frac{1}{2} \right) e^{-\frac{1}{2}t} \frac{1}{t}\, dt.$$

Using this, we obtain

$$\begin{aligned}
\theta(1/2) &= (\theta(1/2) - \theta(1)) + \theta(1) \\
&= \int_0^\infty \left(\frac{e^{-\frac{1}{2}t}}{t} - \frac{1}{e^t - 1} \right) \frac{1}{t}\, dt + \int_0^\infty \left(\frac{1}{e^t - 1} - \frac{1}{t} + \frac{1}{2} \right) e^{-t} \frac{1}{t}\, dt \\
&= \int_0^\infty \left(\frac{e^{-\frac{1}{2}t} - e^{-t}}{t} - \frac{1}{2} e^{-t} \right) \frac{1}{t}\, dt = \int_0^\infty - \frac{d}{dt} \left(\frac{e^{-\frac{1}{2}t} - e^{-t}}{t} \right) - \frac{e^{-\frac{1}{2}t} - e^{-t}}{2t}\, dt.
\end{aligned}$$

Applying (3) we obtain the desired result. ■

Proof of Theorem 1: We have to show that $\varphi(x) = \theta(x)$. By (4), $\theta(x) - \theta(x+1) = \varphi(x) - \varphi(x+1)$. Applying this to $x, x+1, \ldots, x+n-1$ and summing these equations, we see that $\theta(x) - \theta(x+n) = \varphi(x) - \varphi(x+n)$. Since $\lim_{n\to\infty}\theta(x+n) = 0$, we immediately obtain

$$\theta(x) = \varphi(x) - \lim_{n\to\infty}\varphi(x+n) =: \varphi(x) - h(x). \tag{5}$$

Next we show that the function h is decreasing. If $0 \le y \le x$ and $0 \le p \le 1$ then

$$\begin{aligned}
&\sqrt{x+n}\, p^{x+n} - \sqrt{y+n}\, p^{y+n} \\
&\qquad \le \sqrt{x+n}\, p^{y+n} - \sqrt{y+n}\, p^{y+n} \le \left(\sqrt{x+n} - \sqrt{y+n}\right) p
\end{aligned}$$

for all $n \ge 1$. Noting that $0 \le te^{1-t} \le 1 (t \ge 0)$ and using the definition of φ, we conclude that

$$e^{\varphi(x+n)} - e^{\varphi(y+n)} \le \left(\sqrt{x+n} - \sqrt{y+n}\right) e^{\varphi(1)}.$$

Taking the limit as $n \to \infty$, we obtain $e^{h(x)} - e^{h(y)} \le 0$, i.e., $h(x) \le h(y)$. Since the function h is also periodic with period 1, it must be constant. Applying (5) and Lemma 3, we obtain that $h(x) = h(\frac{1}{2}) = 0$ for all $x > 0$. ■

REFERENCES

1. G. Marsaglia and J. C. W. Marsaglia, A new derivation of Stirling's asymptotic approximation to $n!$, this MONTHLY **97** (1990), 826–829.
2. G. Pólya and G. Szegö, *Problems and Theorems in Analysis*, Springer-Verlag, Berlin-Heidelberg-New York, 1972.
3. E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, (Fourth Edition), Cambridge, 1927.

Department of Mathematics, Technical University of Dresden, Mommsenstr. 13, 01062 Dresden, Germany
sasvari@math.tu-dresden.de

A Simple Proof of Rankin's Campanological Theorem

Richard G. Swan

Change ringing is the traditional English method of ringing church bells. The basic idea is to ring a set of bells in all possible orders (the changes) with no repetition until the initial position recurs. The sequence of changes is usually grouped into blocks, known as 'leads,' of a standard form, and one considers the sequence consisting of the last change in each lead (the 'lead ends'). Each lead end is obtained from the previous one by a permutation depending on the type of lead, and one tries to choose the sequence of leads so that all possible changes occur.

The mathematical problem involved in doing this can be formulated more generally as follows. Given a finite group G with a set of generators, E, one attempts to enumerate the elements of G as $x_1, \ldots, x_n$ (with $n = |G|$) in such a way that for each $i, x_{i+1} = x_i e_i$ for some e_i in E (including $x_1 = x_n e_n$). Many explicit solutions have been given in particular cases, often by quite ingenious methods **[5]**, **[6]**, but few general results seem to be known about the possibility of constructing such a sequence. Aside from the obvious requirement that E generates G, the only necessary condition known to me is a theorem of Rankin **[4]**, which generalizes an earlier result of W. H. Thompson for a special case. This theorem asserts that if $E = \{a, b\}$ has at most 2 elements and if $c = ab^{-1}$ has odd order, then $|G\colon \langle a\rangle|$ and $|G\colon \langle b\rangle|$ must be odd. In fact, Rankin proved a more general result in which E is not required to generate G.

By a *cyclically ordered set* I mean a sequence $x_1, \ldots, x_n$ of distinct elements, two such sequences $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_m)$ being regarded as the same if they differ by a cyclic permutation, i.e., $m = n$ and $y_i = x_{i+k}$ for some fixed k (indices being taken mod n). If G is a finite group and E is a subset of G, an *E-cycle* in G is a cyclically ordered subset $x_1, \ldots, x_n$ of G such that the ratios $x_i^{-1}x_{i+1}$ all lie in E (including $x_n^{-1}x_1$).

Theorem 1 [4]. *Let $E = \{a, b\}$ be a subset of a finite group G. Suppose that G has a partition into r disjoint E-cycles. If $c = ab^{-1}$ has odd order, then $r \equiv |G\colon \langle a\rangle| \equiv |G\colon \langle b\rangle| \bmod 2$.*

Applications to change ringing may be found in **[4]**. Some history of Thompson's work is given in **[2]** and **[1]**. Our objective is to give a very simple proof of the theorem. With no more effort we can actually prove a somewhat more general result. Let X be a finite set and let E be a set of permutations of X. An *E-cycle in X* is a cyclically ordered subset $x_1, \ldots, x_n$ of X such that for each $i = 1, \ldots, n$, we have $x_{i+1} = \alpha_i(x_i)$ for some α_i in E. As always, the indices are taken mod n so that $x_1 = \alpha_n(x_n)$ also.

Theorem 2. *Let $E = \{\alpha, \beta\}$ where α and β are permutations of a finite set X having k and l cycles, respectively. Suppose X has a partition into r disjoint E-cycles. If $\gamma = \beta^{-1}\alpha$ has odd order, then $r \equiv k \equiv l \bmod 2$.*

Theorem 1 is an immediate consequence of Theorem 2. We let $X = G$ and define α and β to be right multiplication by a and b, i.e., $\alpha(x) = xa$ and $\beta(x) = xb$. Then $\gamma(x) = xc$ so γ has the same order as c and the cycles of α and β are just the left cosets of the subgroups $\langle a \rangle$ and $\langle b \rangle$. Therefore $k = |G\colon \langle a \rangle|$ and $l = |G\colon \langle b \rangle|$.

Remark. There are two obvious partitions of X into E-cycles, namely the cycles of α and the cycles of β. The point of Theorem 2 is thus that the parity of r is the same for all partitions into E-cycles if γ has odd order.

For the proof, observe that there is a 1-1 correspondence between partitions of X into disjoint cyclic subsets and permutations π of X, the cyclic subsets being the cycles of π. These cycles are E-cycles if and only if for each x in X we have $\pi(x) = \alpha_x(x)$ for some α_x in E. The parity of the number of cycles is related to the sign of π by the following fact.

Lemma 3 [3, App. A]. *Let π be a permutation of n elements having r cycles. Then* $\operatorname{sgn}(\pi) = (-1)^{n+r}$.

In fact, if π has p even cycles and q odd cycles, then $\operatorname{sgn}(\pi) = (-1)^p$, $r = p + q$, and $n \equiv q \bmod 2$.

In the situation of Theorem 2, let $P = \{x \in X | \pi(x) = \alpha(x)\}$ and $Q = \{x \in X | \pi(x) = \beta(x)\}$. Then $X = P \cup Q$. Let $\tau = \beta^{-1}\pi$. Then τ acts as the identity on Q, and $P - Q = X - Q$ is stable under τ, which clearly agrees with γ on it. So $\tau|P = \gamma|P$ and $\tau|Q = 1$. Therefore τ has odd order since γ does, so $\operatorname{sgn}(\tau) = 1$. It follows that $\operatorname{sgn}(\pi) = \operatorname{sgn}(\beta)$ and Lemma 3 shows that $r \equiv l \bmod 2$. Similarly, $r \equiv k \bmod 2$. ■

Remark. We can also get some information if γ is not assumed to have odd order. Note that $P \cap Q = F$, the set of fixed points of γ. Since P and Q are stable under γ, they are determined by their images $\overline{P} = P/\langle \gamma \rangle$ and $\overline{Q} = Q/\langle \gamma \rangle$ in $\overline{X} = X/\langle \gamma \rangle$. Lemma 3 shows that $\operatorname{sgn}(\tau) = (-1)^d$, where $d = |P| + |\overline{P}|$. Since $\operatorname{sgn}(\pi) = \operatorname{sgn}(\beta)\operatorname{sgn}(\tau)$, we see that in all cases $r \equiv l + |P| + |\overline{P}| \bmod 2$. Similarly, $r \equiv k + |Q| + |\overline{Q}| \bmod 2$. In the situation of Theorem 1, $|P| = |\langle c \rangle| \cdot |\overline{P}|$ so if c has even order we get $r \equiv l + |\overline{P}| \bmod 2$ and similarly $r \equiv k + |\overline{Q}| \bmod 2$. It is also easy to see that, in the situation of Theorem 2, the possible partitions of X into E-cycles are in $1-1$ correspondence with subsets $\overline{P}$ of $\overline{X}$ that contain $\overline{F} = F$: We let $\overline{Q} = (\overline{X} - \overline{P}) \cup F$, let P and Q be the inverse images of $\overline{P}$ and $\overline{Q}$ in X, and define π to be α on P and β on Q.

Examples. The fact that P and Q are stable under γ was Thompson's key observation on which all proofs of the theorem are based. It has no analogue if E has more than 2 elements and it is not at all clear whether there is any analogue to Rankin's theorem for this case. One might guess that a similar conclusion holds if $E = \{e_1, \ldots, e_k\}$ and we assume that all the elements $e_i e_j^{-1}$ have odd order. However, this is not the case. We give two examples, one having all indices $|G\colon \langle x \rangle|$ odd for $x \in E$ and one having all these indices even. Following the usual convention [2], an E-cycle $x_1, \ldots, x_k$ having $x_{i+1} = x_i e_i$ with e_i in E and with $x_1 = 1$ is described by writing the word $e_1 e_2 \cdots e_k = 1$. If $x_1 \neq 1$, I write $x_1 \circ e_1 e_2 \cdots e_k = x_1$ instead.

(1) Let $G = S_3$ (the symmetric group) and let $E = \{a, b, c\}$ be the set of elements of order 2. It is well known [2] that there is a partition with $r = 1$, namely, $(ab)^3 = 1$. But there is also one with $r = 2$, namely, $abac = 1$ and $b \circ a^2 = b$.

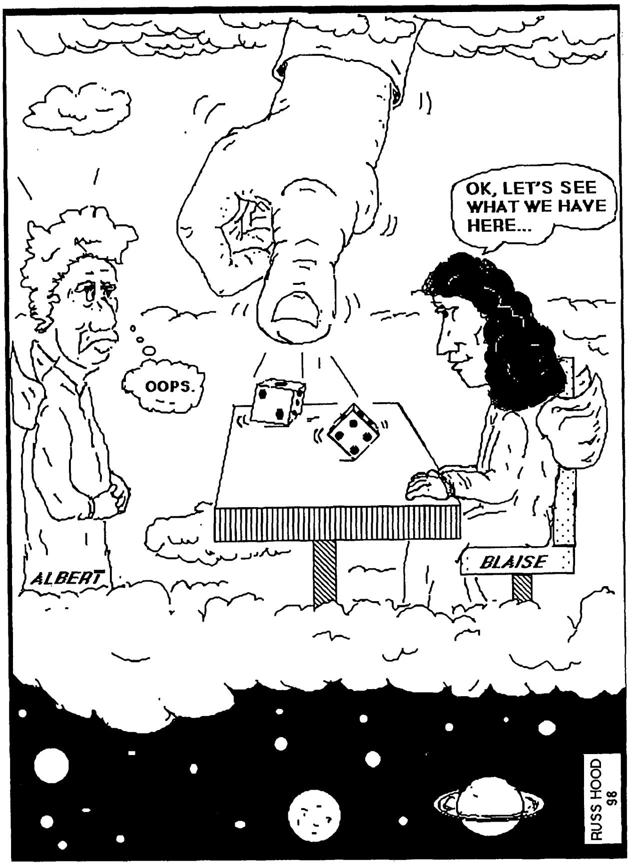
(2) Let $G = A_4$ (the alternating group) and let $E = \{a, b, c\}$ with $a = (12)(34)$, $b = (123)$, and $c = (234)$. The coset decompositions for $\langle a \rangle$, $\langle b \rangle$, and $\langle c \rangle$ all have r even but there is also a partition with $r = 1$, namely, $(c^2ac^2b)^2 = 1$.

ACKNOWLEDGMENT. I thank A. T. White for useful comments on an earlier version of this paper.

REFERENCES

1. D. J. Dickinson, On Fletcher's paper "Campanological Groups," *Amer. Math. Monthly* **64** (1957) 331–332.
2. T. J. Fletcher, Campanological groups, *Amer. Math. Monthly* **63** (1956) 619–626.
3. A. G. Kurosh, *The Theory of Groups*, vol. 1, Chelsea, New York, 1955.
4. R. A. Rankin, A campanological problem in group theory, *Proc. Camb. Phil. Soc.* **44** (1948) 17–25.
5. A. T. White, Ringing the cosets, *Amer. Math. Monthly* **94** (1987) 721–746.
6. A. T. White, Fabian Stedman: The first group theorist?, *Amer. Math. Monthly* **103** (1996) 771–778.

The University of Chicago, 5734 University Avenue, Chicago, IL 60637-1538
swan@math.uchicago.edu



Contributed by Russ Hood. Rio Linda, CA

UNSOLVED PROBLEMS

Edited by **Richard Nowakowski**

In this department the MONTHLY presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial or related results. Typescripts should be sent to Richard Nowakowski, Department of Mathematics and Statistics, Dalhousie University, Halifax NS, Canada B3H 3J5, rjn@cs.dal.ca

Beggar My Neighbour

Marc M. Paulhus

> "What do you play, boy?" asked Estella of myself, with the greatest disdain.
> "Nothing but beggar my neighbour, miss."
> "Beggar him," said Miss Havisham to Estella. So we sat down to cards.
> ...
> I played the game to an end with Estella, and she beggared me. She threw the cards down on the table when she had won them all, as if she despised them for having been won of me.
>
> *Charles Dickens, Great Expectations (1860)*

Many readers can probably recall childhood hours that were whiled away playing a simple card game as Pip and Estella did, although we hope your luck was better than poor Pip's.

Dickens knew the game as "Beggar my Neighbour" but "Strip Jack Naked" is also common. It has been suggested that it is also called "Beat your Neighbour out of Doors" and it may even be the same game as "Knave out of Doors" as mentioned in John Haywood's "A Woman Killed with Kindness" (1607) **[1]**. Similar games include an Italian version called "Camicia" which is played with a different deck and "Egyptian Ratscrew" (or "Egyptian War" or "Bloodystump" or "Egyptian Rhapsody"), which has added elements of speed and violence **[2]**.

You may recall that the game often went on for a very long time, with first one person accumulating a lot of cards, then another, so that bedtime or boredom arrived before a winner could be decided. A question that has been asked, perhaps as long as the game has been played, but certainly by John Conway, is: can the game go on forever? On this topic, Conway wrote:

> This was one of my "anti-Hilbert problems". With the standard pack of 52 cards, I just don't know, and it's not for want of trying. What I do know is that there are cycles in some smaller packs ...
>
> My guess is that there ARE cycles (after all, their existence with small packs shows that there's no magic reason why there shouldn't be), and that a clever enough computer search would probably find one. I've played some games that went on very long and seemed to be cyclic, but they've always ended. It's not only true that "of course, any mistake produces gravitational waves"—it also beggars either me or my neighbor.

How is it played? An ordinary deck of 52 cards is divided as equally as possible among the players, who hold their respective shares face down. Players in rotation take one card from the top of their stack and place it face up on a stack in the center of the table. Play continues until a court card (J, Q, K, or A) is played, whereupon the next player is required to contribute respectively 1, 2, 3, or 4 cards to the central stack. If one of these 1, 2, 3, or 4 cards is a court card, then the player stops contributing, and the onus to supply the appropriate number of cards to the central stack passes to the next player. If none of the 1, 2, 3, or 4 cards is a court card, then the last player to play a court card collects the whole of the central stack, turns it over, and adds it to the underside of his own stack. This player then starts play again by turning the top card of his stack and placing it face up in the center of the table. Any player who runs out of cards, drops out of the game. If this happens during the contribution process, then the obligation to complete the contribution passes to the next player. The winner is the player who accumulates the whole deck.

As Conway says, there are cycles with small decks. Here is one:

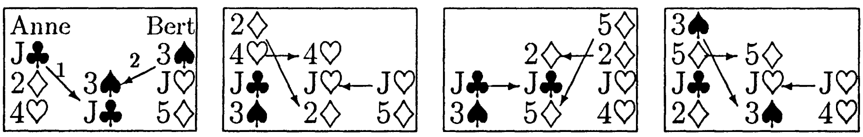


Figure 1. A never-ending mini-game of Beggar-My-Neighbor

and this extends to more players, if Carla, Dan, ... each start with a hand similar to Bert's. A slightly less trivial example, in which a court card does change hands, is the following in which the tops of the hands are on the left. Anne starts with the 9♣.

Anne 9 ♣ 3 ♦ J ♥ 8 ♠

Bert K ♥ 7 ♣ J ♠ 4 ♦

where the King could as well have been a Queen or an Ace.

A much less trivial unending game is

Anne 9♣ 4♣ J♣ 6♣ 3♣ 10♣ Q♣ 2♣ A♣ 5♣ 7♣ K♣ A♦ 8♣

Bert K♦ 3♦ 9♦ J♦ 8♦ 4♦ Q♦ 7♦ 2♦ 10♦ 6♦ 5♦ 2♠ 2♥

You may wonder what those extra 2♠ and 2♥ are doing in there. An exhaustive search on two-player games with half a deck (just two 13-card suits) revealed no cycles at all. But if we add in or remove just two common cards, then there are cycles!

A *deal* is a legal starting position, i.e., the cards are distributed as equally as possible between the players, who are ordered, and one player is designated to start. A *move* in a game is when the cards in the middle are collected by one of the players. Periods and preperiods are reported in terms of moves. A *position* is an ordered set of hands and a bullet to indicate who is to play. To save space, positions are reported without suits and with 0 for common cards, 1 for the Jack, 2 for the Queen, etc...

Given a deck of cards C, which is not necessarily the standard deck, we can construct a directed graph $D_n(C)$ where n is the number of players. Each node in the graph is a position and has one outarrow which points to the position which would result after one move in the game. Some nodes have invalence zero. Naturally, we are most interested in the portion of $D_n(C)$ that is reachable from a deal, which we will call $D'_n(C)$. If the deck C has a unending game then it appears as a cycle in $D'_n(C)$. Note that, with the exception of positions with zero invalence, $D_n(C)$ appears as a subgraph of $D_{n+1}(C)$.

Table 1 shows the results on some exhaustive searches on two-player deals. The last column reports that there is at least one cycle with 2J, 2Q, 2K, 2A, and 20 common cards. This cycle was found by random sampling.

TABLE 1 Results of an exhaustive search of two-player games on decks of cards with 2J, 2Q, 2K, 2A, and n common cards. The second row reports the number of deals with cycles and the final row reports the probability the second player wins.

n	0	2	4	6	8	10	12	14	16	18	20
	0	0	0	1260	11928	4308	0	0	36	0	≥ 1
	1	.60	.533	.514	.508	.506	.505	.504	.504	.503	?

Table 2 shows the 36 deals that never end when two players play with a deck of 2J, 2Q, 2K, 2A, and 16 common cards. Remarkably, there is is essentially only one cycle, which all these games eventually enter. That cycle has a period of 11 moves.

If C is a full deck of cards, does $D'_2(C)$ have a cycle? We leave this question unanswered except to say that we have been unable to find one in 3.2 billion randomly chosen deals. Of course we have searched only a very small portion of the $52!/(36!(4!)^4) = 653{,}534{,}134{,}886{,}878{,}245{,}000$ starting positions. Note that there are an equal number of terminal positions. The longest game we have found is

$$\begin{array}{l} \bullet 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,3\,4\,2\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0 \\ 0\,1\,2\,2\,3\,0\,0\,0\,3\,0\,0\,0\,0\,1\,3\,0\,0\,2\,4\,0\,4\,0\,1\,4\,0\,0 \end{array}$$

which requires that 4791 cards be played before terminating. The average game plays about 254 cards before terminating and there appears to be a nearly nonexistent advantage to going second.

We also played 1 billion random deals with 4 players and were unable to find a cycle. The average game plays about 364 cards before terminating and you should prefer to go 4th rather than 3rd, and 3rd rather than 2nd, and 2nd rather than 1st, although position makes only a small difference to your chances of winning.

Let's look more closely at the graph $D_2(C)$. It consists of a large number of connected components. Every terminal position is a member of a tree component and every tree component contains exactly one terminal position. If there are cycles in $D_2(C)$ then the average size of a tree is less than 52. Hence, by randomly sampling terminal nodes and determining the size of their corresponding trees, you can try to establish statistical evidence to support the existence of cycles. As you might expect, when C is the standard deck, the results are inconclusive.

For the record, the Italian game of "Camicia" is played with a 40 card deck with 12 court cards (4 each of values 1, 2, and 3). Playing a billion random deals of Camicia also failed to produce any cycles.

TABLE 2 The 36 deals that cycle when C is a deck with 2J, 2Q, 2K, 2A, and 16 common cards. As an exercise, the reader may want to draw the subgraph of $D_2(C)$ that is reachable from these deals to see that they all reach essentially the same cycle.

Deal (preperiod)	Deal (preperiod)	Deal (preperiod)
●100020000320 (2) 001000004403	●100040342030 (2) 002100000000	●100000000204 (2) 042030100300
●010000000002 (2) 400420301003	●010000044030 (2) 002000032010	●001000040342 (2) 300021000000
●201000004034 (2) 030000210000	●021000000000 (1) 004034203010	●0301000000200 (2) 020310000044
●030140000200 (2) 020031000004	●320100000040 (2) 403000002100	●0203100000444 (2) 301000002000
●003010400002 (2) 002040310000	●2030100400000 (3) 00200403100	●000021000000 (2) 100004034203
●300021000000 (2) 010000403420	●020031000004 (2) 301400002000	●420301003000 (0) 000000020410
●020301000400 (3) 000020004031	●003201000000 (2) 440300000021	●044030100000 (2) 003201200000
●000320120000 (2) 044030100000	●002030130004 (3) 000002010040	●00204310000 (2) 030104000020
●000002010040 (3) 020301300040	●000032010200 (2) 004403001000	●400420301003 (2) 100000000020
●004403001000 (2) 000320102000	●000003201002 (2) 000440300010	●403000002100 (2) 201000000403
●002000032010 (2) 100000440300	●004034203010 (1) 210000000000	●000440300010 (1) 000032010020
●040300000021 (2) 320100000004	●000020004031 (3) 203010004000	●440300000021 (2) 032010000000

ACKNOWLEDGMENT. We thank Richard K. Guy for his help with this work.

Added in Proof. Since this paper was written we have learned that Michael Kleber has independently established the results in Table 1. He also discovered a longer full-deck game, namely

- 0001200030400004010300023 0
- 0100000000000412400003 0002

which requires 5790 cards (805 moves) to terminate.

REFERENCES

1. David Parlett, *Oxford History of Card Games*, Oxford Univ Press, 1990.
2. John McLeod's webpage: www.pagat.com.

University of Calgary, Calgary, Alberta, Canada, T2N 1N4.
paulhusm@math.ucalgary.ca

PROBLEMS AND SOLUTIONS

Edited by **Gerald A. Edgar, Daniel H. Ullman,** and **Douglas B. West**
with the collaboration of Paul T. Bateman, Mario Benedicty, Paul Bracken, Duane M. Broline, Ezra A. Brown, Richard T. Bumby, Glenn G. Chappell, Randall Dougherty, Roger B. Eggleton, Ira M. Gessel, Bart Goddard, Jerrold R. Griggs, Douglas A. Hensley, Richard Holzsager, John R. Isbell, Robert Israel, Kiran S. Kedlaya, Murray S. Klamkin, Fred Kochman, Frederick W. Luttmann, Vania Mascioni, Frank B. Miles, Richard Pfiefer, Leonard Smiley, John Henry Steelman, Kenneth Stolarsky, Richard Stong, Charles Vanden Eynden, and William E. Watkins.

Proposed problems and solutions should be sent in duplicate to the MONTHLY *problems address on the inside front cover. Submitted problems should include solutions and relevant references. Submitted solutions should arrive at that address before July 31, 1999; Additional information, such as generalizations and references, is welcome. The problem number and the solver's name and address should appear on each solution. An acknowledgement will be sent only if a mailing label is provided. An asterisk (*) after the number of a problem or a part of a problem indicates that no solution is currently available.*

PROBLEMS

10711. *Proposed by Florian Luca, Universität Bielefeld, Bielefeld, Germany.* A natural number is *perfect* if it is the sum of its proper divisors. Prove that two consecutive numbers cannot both be perfect.

10712. *Proposed by Paul Deiermann, Lindenwood University, St. Charles, MO, and Rick Mabry, Louisiana State University, Shreveport, LA.* Let $f(x)$ and $g(y)$ be twice continuously differentiable functions defined in a neighborhood of 0, and assume that $f(0) = 1$, $g(0) = f'(0) = g'(0) = 0$, $f''(0) < 0$, and $g''(0) > 0$.
(**a**) For sufficiently small $r > 0$, show that the curves $x = g(y)$ and $y = rf(x/r)$ have a common point (x_r, y_r) in the first quadrant with the property that, if (x, y) is any other common point, then $x_r < x$.
(**b**) Let $(t_r, 0)$ denote the x-intercept of the line passing through $(0, r)$ and (x_r, y_r). Show that $\lim_{r\to 0^+} t_r$ exists, and evaluate it.
(**c**) Is the continuity of f'' and g'' a necessary condition for $\lim_{r\to 0^+} t_r$ to exist?

10713. *Proposed by Juan-Bosco Romero Márquez, Universidad de Valladolid, Valladolid, Spain.* Given a triangle with angles $A \geq B \geq C$, let a, b, and c be the lengths of the corresponding opposite sides, let r be the radius of the inscribed circle, and let R be the radius of the circumscribed circle. Show that A is acute if and only if $R + r < (b + c)/2$.

10714. *Proposed by Jet Wimp, Drexel University, Philadelphia, PA.* For $a \in (-\pi/2, \pi/2)$, define

$$c_n(t) = \frac{1}{e^{at}\cos a}\left(\frac{d}{da}\right)^n \left(e^{at}\cos a\right)$$

for every nonnegative integer n, so that $c_n(t)$ is a monic polynomial of degree n. Let G_n denote the $(n+1)$-by-$(n+1)$ determinant $\left|c_{j+k}(t)\right|_{j,k=0,1,\ldots,n}$. Evaluate G_n.

10715. *Proposed by Roger Cuculière, Clichy, France.* Choose $u_0 > 1$, and define $u_{n+1} = u_n + \ln u_n$ for $n \in \mathbb{N}$. Find a closed-form expression a_n such that $\lim_{n\to\infty}(u_n - a_n)/n = 0$.

10716. *Proposed by Michael L. Catalano-Johnson and Daniel Loeb, Daniel Wagner Associates, Malvern, PA.* What is the largest cubical present that can be completely wrapped (without cutting) by a unit square of wrapping paper?

10717. *Proposed by Marcin Mazur, University of Chicago, Chicago, IL.* We say that a tetrahedron is *rigid* if it is determined by its volume, the areas of its faces, and the radius of its circumscribed sphere. We say that a tetrahedron is *very rigid* if it is determined just by the areas of its faces and the radius of its circumscribed sphere.
(**a**) Prove that every tetrahedron with faces of equal area is rigid.
(**b**) Prove that a very rigid tetrahedron with faces of equal area is regular.
(**c**)* Is every tetrahedron rigid?
(**d**)* Is every very rigid tetrahedron regular?

SOLUTIONS

Subtracting Square Roots Repeatedly

10568 [1997, 68]. *Proposed by Donald E. Knuth, Stanford University, Stanford, CA.* Let n be a nonnegative integer. The sequence defined by $x_0 = n$ and $x_{k+1} = x_k - \lceil\sqrt{x_k}\rceil$ for $k \geq 0$ converges to 0. Let $f(n)$ be the number of steps required; i.e., $x_{f(n)} = 0$ but $x_{f(n)-1} > 0$. Find a closed form for $f(n)$.

Solution by Denis Constales, University of Gent, Gent, Belgium. Every positive integer n can be written uniquely in the form $p^2 - q$, where p and q are integers satisfying $p \geq 1$ and $0 \leq q \leq 2p - 2$ (take $p = \lceil\sqrt{n}\,\rceil$ and $q = p^2 - n$). We call this *standard form* for n. We obtain the desired formula in terms of these parameters p and q.

Using standard form, let $n' = n - \lceil\sqrt{n}\,\rceil = p^2 - (q + p)$. We distinguish two cases.
Case 1: $p - 1 \leq q \leq 2p - 2$. We rewrite n' as $(p-1)^2 - (q - (p-1))$. Since $q \geq p - 1$, this expresses n' in standard form with $p' = p - 1$ and $q' = q - (p - 1)$ (when $p > 2$).
Case 2: $0 \leq q \leq p - 1$. Now $n' = p^2 - (q + p)$ is standard form for n' with $p' = p$ and $q' = q + p$. The next value $n'' = n' - \lceil\sqrt{n'}\,\rceil = p^2 - (q + 2p)$. Expressed in standard form, this is $n'' = (p-1)^2 - (q + 1)$ (when $p > 2$).

We have applied the transformation once in Case 1 and twice in Case 2. Thus

$$f(p^2 - q) = \begin{cases} 2 + f((p-1)^2 - (q+1)) & \text{if } 0 \leq q \leq p - 2 \\ 1 + f((p-1)^2 - (q - p + 1)) & \text{if } p - 1 \leq q \leq 2p - 2 \end{cases}$$

whenever $p > 2$ and $0 \leq q \leq 2p - 2$. The cases $p \leq 2$ occur for $n \in \{1, 2, 3, 4\}$, where $f(n) = 1, 1, 2, 2$, respectively. With the recurrence, these initial conditions define f. Our closed form is

$$f(p^2 - q) = \begin{cases} 2p - \lfloor\log_2(p+q)\rfloor - 1 & \text{if } 0 \leq q \leq p - 1 \\ 2p - \lfloor\log_2 q\rfloor - 2 & \text{if } p \leq q \leq 2p - 2 \end{cases}$$

for integers p, q such that $1 \leq p$ and $0 \leq q \leq 2p - 2$. Also, we set $f(0) = 0$.

The proof of the formula is immediate by induction, using the recurrence in the three cases $0 \leq q \leq p - 2$, $q = p - 1$, and $p \leq q \leq 2p - 2$. The only simplification needed occurs in the second case, where $\lceil\log_2(2p - 1)\rceil = 1 + \lceil\log_2(p - 1)\rceil$, which follows immediately when $p > 1$.

Editorial comment. Robin J. Chapman and the GCHQ Problems Group expressed $f(n)$ using the single formula $f(n) = \lfloor 4n + 2^{m+3} - 3\rfloor - (m + 2)$, where $m = \lfloor\log_2(\sqrt{n} + 1)\rfloor$.

Solved also by T. Amdeberhan, K. L. Bernstein, R. J. Chapman (U. K.), D. A. Darling, M. N. Deshpande & N. N. Kasturiwale (India), K. Ferguson, R. Holzsager, W. Janous (Austria), F. Kemp, P. G. Kirmser, N. Komanda, Y. Kong, J. H. Lindsey II, W. A. Newcomb, C. R. Pranesachar (India), K. Schilling, J. H. Steelman, D. Trautman, X. Wang, D. Yuen, GCHQ Problems Group (U. K.), Westmont Problems Group, and the proposer.

Graphs without Increasing Paths

10572 [1997, 168]. *Proposed by Richard P. Stanley, Massachusetts Institute of Technology, Cambridge, MA.* Let $f(n)$ be the number of graphs (without loops or multiple edges) on the vertices $1, 2, \ldots, n$ such that no path of length two has vertices i, j, k (in that order) with $i < j < k$. Let $g(n)$ be the total number of subspaces of an n-dimensional vector space over a 2-element field. show that

$$\sum_{n\geq 0} f(n)\frac{x^n}{n!} = e^{-x}\sum_{n\geq 0} g(n)\frac{x^n}{n!}.$$

Solution by Stephen C. Locke, Florida Atlantic University, Boca Raton, FL. Let V be an n-dimensional binary vector space, and let S be a subspace of V. We may take S to be the row space of an $m \times n$ binary matrix M. Furthermore, we may assume that M is row-reduced, so that the leading 1 in each row is the only 1 in its column. We call these entries *pivots*.

Construct a graph G whose vertices are the nonzero columns of M (G is empty when S has dimension 0). For each pivot element $m_{i,j}$, the vertex representing column j is adjacent to the vertex representing column r if $m_{i,r} \neq 0$. Thus all edges consist of a pivot column and a higher-indexed non-pivot column. In particular, G has no path i, j, k with $i < j < k$. Furthermore, the row-reduced matrix M and thus S can be retrieved from G.

If we relabel the vertices of G with $1, 2, \ldots, k$ preserving the order of the original labels, then the new graph is of the type counted by $f(k)$. Thus there are $\sum_{k=0}^{n}\binom{n}{k}f(k)$ such graphs, and the bijection with subspaces yields $g(n) = \sum_{k=0}^{n}\binom{n}{k}f(k)$.

When multiplying power series, the coefficient of $x^n/n!$ in the product of $\sum_{n\geq 0} a_n x^n/n!$ and $\sum_{n\geq 0} b_n x^n/n!$ is $\sum_{k=0}^{n}\binom{n}{k}a_k b_{n-k}$. Thus $\sum_{n\geq 0} g(n)x^n/n! = e^x \sum_{n\geq 0} f(n)x^n/n!$.

Solved also by D. Beckwith, D. Callan, R. J. Chapman (U. K.), J. E. Dawson (Australia), R. Ehrenborg, R. Holzsager, D. E. Knuth, L. Pebody (U. K.), and the proposer.

A Card-Matching Game

10576 [1997, 169]. *Proposed by Donald E. Knuth, Stanford University, Stanford, CA.* Alice and Bill have identical decks of 52 cards. Alice shuffles her deck and deals the cards face up into 26 piles of two cards each. Bill does the same with his deck. If any one of Alice's top cards exactly matches any of Bill's, the matching cards are removed. Play continues until none of the cards on top of Alice's piles matches any of the cards on top of Bill's piles. What is the probability that all 52 pairs of cards will be matched?

Solution by Philip D. Straffin, Beloit College, Beloit, WI. Given a particular deal of $2n$ cards for Alice, let T_n be the number of possible deals for Bill, let W_n be the number of these that succeed (all cards are matched), and let L_n be the number that lose. We compute W_n/T_n.

Given a deal for Alice, a game is specified by Bill's piles: a partition of $\{1, 2, \ldots, 2n\}$ into pairs and a choice for the top card in each pair. After specifying which of Bill's cards is paired with 1 and which is the top card in this pair, there are T_{n-1} ways to complete the deal. Hence $T_n = 2(2n-1)T_{n-1}$.

Now consider L_n. As long as either player has any single-card piles, the game is not lost, since more than half of that player's cards are exposed and there must be a match. Hence when a game is lost, each player retains $k \geq 1$ piles of two cards, and Bill's top cards must be exactly Alice's bottom cards. The piles that were removed were a successful game of size $n-k$. Since the hidden cards in Bill's blocked piles are chosen from Alice's top cards and can be arranged in $k!$ ways, we have $L_n = \sum_{k=1}^{n}\binom{n}{k}k!\,W_{n-k}$.

Since $T_n = W_n + L_n$, we have $T_n = \sum_{k=0}^{n}\binom{n}{k}k!\,W_{n-k}$. Thus

$$nT_{n-1} = n\sum_{k=0}^{n-1}\binom{n-1}{k}k!\,W_{n-1-k} = \sum_{k=0}^{n-1}\binom{n}{k+1}(k+1)!\,W_{n-1-k} = \sum_{k=1}^{n}\binom{n}{k}k!\,W_{n-k} = L_n.$$

With the recurrence for T_n, this yields

$$\frac{L_n}{T_n} = \frac{nT_{n-1}}{2(2n-1)T_{n-1}} = \frac{n}{4n-2},$$

and the probability of matching all cards is $(3n-2)/(4n-2)$. When $n = 26$, this is about 0.745.

Editorial comment. The GCHQ Problems Group considered a generalization: Alice starts with a piles of two cards (doubles) and $c - 2a$ piles of one card (singles), while Bill starts with b doubles and $c-2b$ singles. The probability of failure is $ab/(c(c-1))$, for $c \geq 2$. The proposer observed that his problem is a variation on one that Lewis Carroll recorded in his diary on February 29, 1856. That problem, called Sympathy, was given to him by someone named Pember and remains unsolved. In Sympathy, the cards are dealt into 18 piles of sizes $3, 3, \ldots, 3, 1$ instead of 26 piles of size 2.

Solved also by D. Beckwith, D. Callan, R. J. Chapman (U. K.), D. A. Darling, J. E. Dawson (Australia), R. Ehrenborg, P. Griffin, C. M. Grinstead, V. Hernández & J. Martín (Spain), R. Holzsager, M. A. Javaloyes Victoria (Spain), J. T. Lee, J. H. Lindsey II, J. H. Nieto (Venezuela), L. Pebody (U. K.), B. Peterson, M. A. Prasad (India), A. L. Rocha, W. J. Seaman, D. S. Silver & S. G. Williams, M. Woltermann, N. Zoroa & P. Zoroa (Spain), GCHQ Problems Group (U. K.), and the proposer.

Wilson's Theorem in Disguise

10578 [1997, 270]. *Proposed by Herbert S. Wilf, University of Pennsylvania, Philadelphia, PA.* Consider the sequence $y_2, y_3, \ldots$ defined by the recurrence relation

$$(n+1)(n-2)y_{n+1} = n(n^2-n-1)y_n - (n-1)^3 y_{n-1}$$

and initial conditions $y_2 = y_3 = 1$. Show that y_n is an integer if and only if n is prime.

Solution by Florian Herzig, Perchtoldsdorf, Austria. Let $x_n = ny_n$ for all $n \geq 2$. We have $x_2 = 2$, $x_3 = 3$, and $(n-2)x_{n+1} = (n^2-n-1)x_n - (n-1)^2 x_{n-1}$ for $n \geq 3$, which becomes

$$\frac{x_{n+1} - x_n}{n-1} = (n-1) \cdot \frac{x_n - x_{n-1}}{n-2}.$$

Setting $z_n = (x_{n+1} - x_n)/(n-1)$ yields $z_n = (n-1)!$, and so $x_{n+1} - x_n = (n-1)z_n = (n-1)(n-1)! = n! - (n-1)!$. Hence

$$x_n = x_2 + \sum_{k=2}^{n-1}(x_{k+1} - x_k) = 2 + (n-1)! - 1! = (n-1)! + 1.$$

It follows that $y_n = x_n/n = ((n-1)!+1)/n$. By Wilson's Theorem, $(n-1)!+1$ is divisible by n if and only if n is prime. Hence y_n is an integer if and only if n is prime.

Editorial comment. In some books, Wilson's Theorem is the statement that $(n-1)!+1$ is divisible by n when n is prime. The converse is also well known and is easily established, since $ny_n = (n-1)!+1$ requires that n and $(n-1)!$ be relatively prime.

Walther Janous noted that one might also study sequences of the form $y_n(a) = \frac{n!+a}{n+a}$ for any integer a. He asks whether there are any integers $a > 1$ for which this sequence contains infinitely many integers; either answer suggests other interesting questions.

Solved also by R. Akhlaghi & F. Sami, J. Anglesio (France), M. N. Balachandran (India), R. Barbara (Lebanon), C. Berg (Sweden), E. Brown, M. Burger (Austria), S. Butcher & X. Wang, D. Callan, R. J. Chapman (U. K.), M. P. Chernesky, B. Conolly (U. K.), D. A. Darling, J. E. Dawson (Australia), D. Donini (Italy), H. Gauchman, C. Georghiou (Greece), R. Heller, R. Holzsager, T. Jager, W. Janous (Austria), W. Kim (South Korea), R. A. Kopas, J. H. Lindsey II, S. C. Locke, R. Martin (Germany), V. J. Matsko, B. McCabe, J. H. Nieto (Venezuela), R. Padma (India), M. D. Pearce, W. H. Pierce, J. Robertson, R. K. Schwartz, Z. Shan & E. T. H. Wang (Canada), P. Simeonov, A. Sinefakopoulos (Greece), N. C. Singer, A. Stenger, D. C. Terr, A. Tissier (France), J. Van hamme (Belgium), J. H. van Lint (The Netherlands), M. Vowe (Switzerland), GCHQ Problems Group (U. K.), NSA Problems Group, and the proposer.

Möbius and Riemann

10582 [1997, 270]. *Proposed by Peter Lindqvist and Kristian Seip, Norwegian University of Science and Technology, Trondheim, Norway.* Let $\mu(n)$ denote the Möbius function of number theory, and let $\zeta(s)$ denote the Riemann zeta function. Prove that

$$\zeta(s)\sum_{m=1}^{N}\sum_{n=1}^{N}\frac{(\gcd(m,n))^s}{(mn)^s}\mu(m)\mu(n) = 1+\sum_{j=2}^{\infty}\frac{1}{j^s}\left(\sum_{\substack{n|j\\ n>N}}\mu(n)\right)^2$$

when $s > 1$.

Solution by David M. Bradley, University of Maine, Orono, ME. We use the well-known fact that $\sum_{n|j}\mu(n) = 0$ for $j \geq 2$. We compute

$$\begin{aligned}
1+\sum_{j=2}^{\infty}\frac{1}{j^s}\left(\sum_{\substack{n|j\\ n>N}}\mu(n)\right)^2 &= 1+\sum_{j=2}^{\infty}\frac{1}{j^s}\left(\sum_{n|j}\mu(n)-\sum_{\substack{n|j\\ n\le N}}\mu(n)\right)^2\\
&= 1+\sum_{j=2}^{\infty}\frac{1}{j^s}\left(\sum_{\substack{n|j\\ n\le N}}\mu(n)\right)^2 = 1+\sum_{j=2}^{\infty}\frac{1}{j^s}\left(\sum_{\substack{m|j\\ m\le N}}\mu(m)\right)\left(\sum_{\substack{n|j\\ n\le N}}\mu(n)\right).
\end{aligned}$$

In the inner sums, m and n both divide j if and only if $\operatorname{lcm}(m,n)|j$. Writing $j = k\cdot\operatorname{lcm}(m,n)$ and interchanging the order of summation yields

$$1+\sum_{j=2}^{\infty}\frac{1}{j^s}\left(\sum_{\substack{n|j\\ n>N}}\mu(n)\right)^2 = \sum_{m=1}^{N}\sum_{n=1}^{N}\frac{\mu(m)\mu(n)}{(\operatorname{lcm}(m,n))^s}\sum_{k=1}^{\infty}\frac{1}{k^s}.$$

Since $\operatorname{lcm}(m,n) = mn/\gcd(m,n)$, the result follows.

Editorial comment. The proposers' solution was quite different. They introduced the functions $f(x) = \sum_{k=1}^{\infty}\sin(kx)/k^s$ and $f_N(x) = \sum_{n=1}^{N}(\mu(n)/n^s)f(nx)$. For $N \in \mathbb{N}$, we have $f_N(x) \to \sin x$ as $s \to \infty$, so it is natural to compute the $L_2(0,\pi)$ norm of the "error" $\sin x - f_N(x)$. Doing this in two different ways yields the result.

Solved also by M. N. Balachandran (India), D. Callan, R. J. Chapman (U. K.), R. Holzsager, J. H. Lindsey II, R. Padma (India), P. Simeonov, NSA Problems Group, and the proposers.

Catalan and Hankel

10585 [1997, 361]. *Proposed by Alta Kellogg, Ormond Beach, FL.* A sequence $a_0, a_1, \ldots$ of real numbers is called *strictly totally positive* (STP) if every submatrix of the Hankel matrix $(a_{i+j})_{i,j\geq 0}$ has positive determinant.

(a) Show that the sequence $C_0, C_1, \ldots$ of Catalan numbers, defined by $C_n = \frac{1}{n+1}\binom{2n}{n}$, is STP.

(b) Show that the sequence of Catalan numbers is minimal in the following sense: If $a_0, a_1, a_2, \ldots$ is an STP sequence of positive integers with $a_n \leq C_n$ for every n, then $a_n = C_n$ for every n.

Solution to part **(a)** *by David Callan, Madison, WI.* Let C be the matrix (C_{i+j}). For sets of indices $\mathbf{u} = \{u_1 < \cdots < u_n\}$ and $\mathbf{v} = \{v_1 < \cdots < v_n\}$, let $C[\mathbf{u}|\mathbf{v}]$ denote the submatrix of C with rows indexed by $\mathbf{u}$ and columns indexed by $\mathbf{v}$. Recall that the Catalan number C_k is the number of Dyck paths ("mountain ranges") of length $2k$. (A Dyck path consists of northeast and southeast steps, starts on the x axis, ends on the x axis, and never falls below

the x axis.) We claim that $\det C[\mathbf{u}|\mathbf{v}]$ is the number of n-tuples of pairwise nonintersecting Dyck paths in which the ith path extends from $(-2u_i, 0)$ to $(2v_i, 0)$.

To see this, let $S = S(\mathbf{u}, \mathbf{v})$ denote the set of all n-tuples of Dyck paths such that the ith path extends from $(-2u_i, 0)$ to $(2v_{\phi(i)}, 0)$ and ϕ is a permutation of $\{1, \dots, n\}$. To each such n-tuple, assign the weight $\operatorname{sgn} \phi$. From the definition of $\det C[\mathbf{u}|\mathbf{v}]$ as a sum of signed products, it is immediate that $\det C[\mathbf{u}|\mathbf{v}]$ is the sum of the weights of the n-tuples in S. The n-tuples that have intersections cancel via a sign-reversing involution: on the lexicographically lowest-indexed pair of intersecting paths, locate the first point of intersection and switch the tails of these two paths after this intersection point. This changes the sign of the associated permutation ϕ and is an involution.

Thus only the nonintersecting n-tuples contribute to the sum, and their weight is 1, since avoiding intersections forces ϕ to be the identity. This establishes the claim.

It follows that $\det C[\mathbf{u}|\mathbf{v}]$ is positive, since there is always the nonintersecting n-tuple consisting of paths in which all northeast steps precede all southeast steps.

Solution to part **(b)** *by Robin J. Chapman, University of Exeter, Exeter, U. K.* In the cases in which $\mathbf{u} = \mathbf{v} = \{0, 1, \dots, n-1\}$ or $\mathbf{u} = \{0, 1, \dots, n-1\}$ and $\mathbf{v} = \{1, 2, \dots, n\}$, there is only one nonintersecting n-tuple of paths. Therefore, the matrices $H_{n-1} = (C_{i+j})_{i,j=0}^{n-1}$ and $H'_{n-1} = (C_{i+j+1})_{i,j=0}^{n-1}$ both have determinant 1.

Now suppose that (a_{i+j}) is an STP sequence of positive integers with $a_n \le C_n$ for every n. Since $C_0 = C_1 = 1$, we have $a_0 = a_1 = 1$. For an inductive proof, we suppose that $a_i = C_i$ for $i < n$.

When $n = 2m$ is even, let $A = (a_{i+j})_{i,j=0}^{m}$. Expanding along the bottom row yields

$$1 - \det(A) = \det(H_m) - \det(A) = (C_n - a_n)\det(H_{m-1}) = C_n - a_n \ge 0.$$

Thus $\det(A) \le 1$, but by hypothesis $\det(A) \ge 1$. Hence $\det(A) = 1$ and $a_n = C_n$.

When $n = 2m + 1$ is odd, the argument is similar, setting $A = (a_{i+j+1})_{i,j=0}^{m}$ and considering H'_m rather than H_m.

Editorial comment. The interpretation of a minor of a Hankel determinant in terms of nonintersecting Dyck paths is due to X. G. Viennot, *Une théorie combinatoire des polynômes orthogonaux généraux* (lecture notes), Université du Québec à Montréal, 1983. The Catalan numbers are essentially moments of the Chebyshev polynomials of the second kind.

Another evaluation of the determinant $\det(C_{i+j})_{i,j=0}^{n-1}$ appears in C. Radoux, Nombres de Catalan généralisés, *Bull. Belg. Math. Soc.* 4 (1997) 289–292.

Solved also by the proposer.

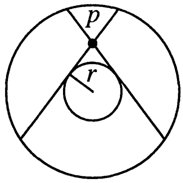
A Random Distance

10592 [1997, 456]. *Proposed by Roger Pinkham, Hoboken, NJ.* Three points are selected independently and at random in a disk of radius one. What is the average distance of the third from the line determined by the first two?

Solution by Richard Holzsager, American University, Washington, DC. Let f be the probability density function of the distance r from the chord through the first two points to the origin and let $\mu = \mu(r)$ be the mean distance from a random point in the circle to a chord at distance r from the origin. Then $\int_0^1 f(r)\mu(r)\,dr$ gives the expected value we want.

We can find f by calculating the probability of a chord being at distance greater than r and differentiating. Let p be the first point, q the second point, and R the distance of the chord through them from the origin. Then $f(r) = -(d/dr)\,P(R > r)$. To find $P(R > r)$, first consider a point p at distance s from the origin. The point q determines a chord at distance greater than r if q is between the two tangents from p to the circle $x^2 + y^2 = r^2$. Thus, given p, the probability that the chord pq is at distance greater than r is $1/\pi$ times

the area of the butterfly-shaped region between the tangents, as shown in the figure at right. To compute this area, consider a variable chord in this region through p.



This chord is divided by p into two segments, one in each wing of the butterfly. Using the polar integral, we can get the area of one of the wings by integrating half the square of the length of one of these segments. It turns out, however, to be convenient to handle both at the same time, rather than finding one area and doubling.

Let θ be the angle between the chord and the radius through p, which is between $\theta_0 = \sin^{-1}(r/s)$ and $\theta_1 = \pi - \sin^{-1}(r/s)$. The lengths of the two parts of the chord satisfy $l_1+l_2 = 2\sqrt{1-s^2\sin^2\theta}$ and $l_1-l_2 = 2s\cos\theta$. Squaring and adding these equations yields $l_1^2+l_2^2 = 2(1+s^2\cos^2(2\theta))$. The area of the region is then $A = \int_{\theta_2}^{\theta_1} 1+s^2\cos(2\theta)\,d\theta = \pi - 2\sin^{-1}(r/s) - 2r\sqrt{s^2-r^2}$. Given the point p, the probability that the chord is at distance greater than r from the origin is, therefore, A/π. The probability distribution function for s is $g(s) = 2s$, so the overall probability that the chord is at distance greater than r is $P(R > r) = 2\int_r^1 s - (2s/\pi)(\sin^{-1}(r/s) + r\sqrt{s^2-r^2})\,ds$.

Fortunately we do not have to calculate this integral to get its derivative. The Fundamental Theorem and differentiation under the integral sign yield

$$f(r) = 2\left(r - \frac{2r}{\pi}\left(\frac{\pi}{2}\right)\right) - 2\int_r^1 -\frac{2s}{\pi}\left(\frac{1}{\sqrt{s^2-r^2}} + \sqrt{s^2-r^2} - \frac{r^2}{\sqrt{s^2-r^2}}\right) ds$$

$$= \frac{4}{\pi}\left(\sqrt{s^2-r^2} + \frac{1}{3}(s^2-r^2)^{3/2} - r^2\sqrt{s^2-r^2}\right)\Bigg|_r^1 = \frac{16}{3\pi}(1-r^2)^{3/2}.$$

Next, consider a chord C at distance r from the origin. A symmetric pair of points at distance less than r from the diameter parallel to C have average distance r from C, while a symmetric pair at distance $x > r$ from the diameter have average distance x from C. Denote by $A(r)$ the area of the smaller region cut off by C within the circle. The overall mean distance from points in the circle to C is $\mu(r) = (1/\pi)\big((\pi - 2A(r))r + 2\int_r^1 x \cdot 2\sqrt{1-x^2}\,dx\big) = r - (2/\pi)rA(r) + (4/3\pi)(1-r^2)^{3/2}$.

Combining the two results gives

$$\int_0^1 f(r)\mu(r)\,dr = \frac{16}{3\pi}\int_0^1 (1-r^2)^{3/2}\left(r - \frac{2}{\pi}rA(r) + \frac{4}{3\pi}(1-r^2)^{3/2}\right) dr$$

$$= \frac{16}{3\pi}\int_0^1 r(1-r^2)^{3/2}dr - \frac{32}{3\pi^2}\int_0^1 r(1-r^2)^{3/2}A(r)dr + \frac{64}{9\pi^2}\int_0^1 (1-r^2)^3dr.$$

Using integration by parts on the middle integral, this works out to

$$-\frac{16}{15\pi}(1-r^2)^{5/2}\Bigg|_0^1 + \frac{32}{3\pi^2}\left(\frac{1}{5}A(r)(1-r^2)^{5/2}\Bigg|_0^1 + \frac{2}{5}\int_0^1 (1-r^2)^3dr\right)$$

$$+\frac{64}{9\pi^2}\int_0^1 (1-r^2)^3dr$$

$$= \frac{16}{15\pi} - \frac{32}{3\pi^2}\left(\frac{\pi}{10}\right) + \left(\frac{64}{15\pi^2} + \frac{64}{9\pi^2}\right)\left(1 - 1 + \frac{3}{5} - \frac{1}{7}\right) = \frac{8192}{1575\pi^2} \approx 0.527.$$

Solved also by D. Callan, K. McInturff, A. Pechtl (Germany), H. A. Steinberg, R. Stong, GCHQ Problems Group (U. K.), and the proposer.

An Infinite Product

10605 [1997, 567]. *Proposed by Jonathan M. Borwein and C. G. Pinner, Simon Fraser University, Burnaby, BC, Canada.* Let r and m be positive integers and define

$$P_r(m) = \prod_{n \neq m} \frac{n^r - m^r}{n^r + m^r}.$$

(**a**) Show that $P_1(m) = 0$ and that

$$P_3(m) = (-1)^{m+1} \frac{2}{3} (m!)^2 \prod_{n=1}^{m} \frac{n+m}{n^3 + m^3}.$$

(**b**) Show that $P_2(m) = (-1)^{m+1} \pi m / \sinh(\pi m)$ and that, more generally, $P_{2s}(m)$ is given by

$$(-1)^{m+1} \frac{2^\epsilon m\pi}{s} (\sinh m\pi)^{(-1)^s} \prod_{j=1}^{s-1} \left(\cosh\left(2\pi m \sin\left(\frac{j\pi}{2s}\right)\right) - \cos\left(2\pi m \cos\left(\frac{j\pi}{2s}\right)\right)\right)^{(-1)^j}$$

where $\epsilon = (1 + (-1)^s)/2$.

Solution by David Bradley, University of Maine, Orono, ME.

(**a**) First, for the case $r = 1$, the infinite product "diverges" to 0 because of the divergence of the harmonic series. Next consider the case $r = 3$. Let $f(n) = n(n-m)/(n^2 - mn + m^2)$. The product becomes $\prod_{n \neq m} f(n)/f(n+m)$. The product now telescopes, and since $f(n) \to 1$ as $n \to \infty$, it reduces to $f(2m) \prod_{n=1}^{m-1} f(n)$ and then to the given expression.

(**b**) For each positive integer r, define $f_r(x) = \prod_{n \geq 1} (n^r - x^r)/(n^r + x^r)$ when x is not an integer. Then for positive integers s and m, we have

$$P_{2s}(m) = \lim_{x \to m} \frac{m^{2s} + x^{2s}}{m^{2s} - x^{2s}} f_{2s}(x).$$

Let $\omega = \exp(i\pi/s)$ and $y = x \exp(-i\pi/2s)$. Then

$$f_{2s}(x) = \prod_{n \geq 1} \frac{n^{2s} - x^{2s}}{n^{2s} + x^{2s}} = \prod_{n \geq 1} \prod_{k=1}^{2s} \frac{n - x\omega^k}{n - y\omega^k}.$$

Using Gauss's infinite product expansion $\Gamma(1+z) = \prod_{n \geq 1} \left(1 + \frac{1}{n}\right)^z \left(1 + \frac{z}{n}\right)^{-1}$, we obtain

$$f_{2s}(x) = \prod_{k=1}^{2s} \frac{\Gamma(1 - y\omega^k)}{\Gamma(1 - x\omega^k)} = \prod_{k=1}^{s} \frac{\Gamma(1 - y\omega^k)\Gamma(1 + y\omega^k)}{\Gamma(1 - x\omega^k)\Gamma(1 + x\omega^k)}.$$

The reflection formula $\Gamma(1-z)\Gamma(1+z) = \pi z/\sin(\pi z)$ — a consequence of Euler's formula $\prod_{n \geq 1}(1 - z^2/n^2) = \sin(\pi z)/(\pi z)$ — now gives

$$\begin{aligned} f_{2s}(x) &= \prod_{k=1}^{s} \frac{\sin(\pi x\omega^k)\pi y\omega^k}{\pi x\omega^k \sin(\pi y\omega^k)} = e^{-i\pi/2} \prod_{j=1}^{2s} \left(\sin\left(\pi x e^{i\pi j/2s}\right)\right)^{(-1)^j} \\ &= -i(\sin(i\pi x))^{(-1)^s} \sin(-\pi x) \prod_{j=1}^{s-1} \left(\sin\left(\pi x e^{i\pi j/2s}\right) \sin\left(\pi x e^{i\pi(2s-j)/2s}\right)\right)^{(-1)^j} \\ &= i2^\epsilon (i\sinh(\pi x))^{(-1)^s} \sin(\pi x) \prod_{j=1}^{s-1} \left(2\sin\left(\pi x e^{i\pi j/2s}\right) \sin\left(-\pi x e^{-i\pi j/2s}\right)\right)^{(-1)^j} \end{aligned}$$

where $\epsilon = (1 + (-1)^s)/2$. We now use the addition formulæ for the cosine to express a product of two sines as a difference of two cosines and simplify to obtain

$$f_{2s}(x) = 2^\epsilon(\sinh(\pi x))^{(-1)^s} \sin(\pi x) \prod_{j=1}^{s-1} \left(\cosh\big(2\pi x \sin(\frac{\pi j}{2s})\big) - \cos\big(2\pi x \cos(\frac{\pi j}{2s})\big)\right)^{(-1)^j}$$

and hence

$$P_{2s}(m) = \lim_{x\to m} \frac{m^{2s} + x^{2s}}{m^{2s} - x^{2s}} f_{2s}(x) = 2^\epsilon\big(\sinh(\pi m)\big)^{(-1)^s} \frac{2\pi m^{2s}\cos(\pi m)}{-2sm^{2s-1}} \times$$
$$\prod_{j=1}^{s-1}\left(\cosh\big(2\pi m \sin(\frac{\pi j}{2s})\big) - \cos\big(2\pi m \cos(\frac{\pi j}{2s})\big)\right)^{(-1)^j}$$
$$= (-1)^{m+1}\frac{2^\epsilon \pi m}{s}\big(\sinh(\pi m)\big)^{(-1)^s} \prod_{j=1}^{s-1}\left(\cosh\big(2\pi m \sin(\frac{\pi j}{2s})\big) - \cos\big(2\pi m \cos(\frac{\pi j}{2s})\big)\right)^{(-1)^j}$$

as required. Note that this formula gives $P_2(m) = (-1)^{m+1}\pi m/\sinh(\pi m)$ when $s = 1$.

Solved also by R. J. Chapman (U. K.), K.-K. Choi, R. Mortini (France), H.-J. Seiffert (Germany), and the proposers.

Monomial Bounds for Polynomials

10613 [1997, 767]. *Proposed by F. J. Flanigan, San Jose State University, San Jose, CA.* Fix a positive real number ν. Find all polynomials $P(x)$ with nonnegative real coefficients such that

(**a**) $P(0) = 0$, $P(1) = 1$, and $P(x) \le x^\nu$ for all $x \ge 0$.
(**b**) $P(0) = 0$, $P(1) = 1$, and $P(x) \ge x^\nu$ for all $x \ge 0$.

Solution by Roberto Tauraso, Firenze, Italy. Let $P(x) = \sum_{i=m}^n a_i x^i$ with nonnegative real coefficients, $a_m > 0$, and $a_n > 0$. The conditions $P(0) = 0$, $P(1) = 1$ imply immediately that $m \ge 1$ and $\sum_{i=m}^n a_i = 1$.
(**a**) If $P(x) \le x^\nu$ for all $x \ge 0$, then necessarily

$$\lim_{x\to+\infty}\frac{P(x)}{x^\nu} = \lim_{x\to+\infty}\frac{a_n x^n}{x^\nu} \le 1 \qquad \text{and} \qquad \lim_{x\to 0^+}\frac{P(x)}{x^\nu} = \lim_{x\to 0^+}\frac{a_n x^m}{x^\nu} \le 1,$$

which imply $n \le \nu$ and $\nu \le m$, respectively. Hence $m = \nu = n$, and condition (**a**) is satisfied if and only if ν is a positive integer and $P(x) = x^\nu$.
(**b**) If $P(x) \ge x^\nu$ for all $x \ge 0$, then the function $\varphi(x) = P(x) - x^\nu$ is nonnegative, differentiable, and satisfies $\varphi(1) = 0$. Hence φ has a minimum at $x = 1$, so $\varphi'(1) = \left(\sum_{i=m}^n i a_i\right) - \nu = 0$. Thus ν is a convex combination of the integers $m, \ldots, n$.

On the other hand, suppose that a polynomial $P(x) = \sum_{i=m}^n a_i x^i$ has nonnegative real coefficients such that $\sum_{i=m}^n a_i = 1$ and $\sum_{i=m}^n i a_i = \nu$. Then $P(0) = 0$, $P(1) = 1$, and, by the weighted arithmetic-geometric mean inequality, $P(x) = \sum_{i=m}^n a_i x^i \ge x^\nu$ for all $x \ge 0$. Thus condition (**b**) is satisfied if and only if $\nu \ge 1$ and $P(x) = \sum_{i=m}^n a_i x^i$, with $\sum_{i=m}^n a_i = 1$ and $\sum_{i=m}^n i a_i = \nu$.

Editorial comment. Erik I. Verriest provided a generalization to the case in which $P(x)$ is a power series. The results are the same as in the selected solution, except that in part (**b**) the upper limit of summation n may be infinite.

Solved also by P. Alsholm (Denmark), K. F. Andersen (Canada), T. Armstrong, M. Babilonová & J. Kupka (Czech Republik), R. J. Chapman (U. K.), J. H. Lindsey II, A. Nijenhuis, C. Popescu (Belgium), H.-J. Seiffert (Germany), E. I. Verriest, GCHQ Problems Group (U. K.), NSA Problems Group, WMC Problems Group, and the proposer.

10621 [1997, 870]. *Proposed by Harold G. Diamond and Bruce Reznick, University of Illinois, Urbana-Champaign, IL.* Let $F(x)$ denote the Cantor singular function, that is, the unique non-decreasing function on [0, 1] such that, if $x = \sum_{j=1}^{\infty} 2\epsilon_j/3^j$ with $\epsilon_j \in \{0, 1\}$, then $F(x) = \sum_{j=1}^{\infty} \epsilon_j/2^j$. It is clear by symmetry that $\int_0^1 F(x)\,dx = 1/2$. Prove that

$$\int_0^1 (F(x))^2\,dx = \frac{3}{10} \quad \text{and} \quad \int_0^1 (F(x))^3\,dx = \frac{1}{5}.$$

More generally, evaluate $\int_0^1 (F(x))^n\,dx$ for every positive integer n.

Solution I by Kenneth F. Andersen, University of Alberta, Edmonton, Alberta. We prove that

$$\int_0^1 (F(x))^n\,dx = \frac{2}{3(n+1)} \sum_{j=0}^{n} \binom{n+1}{j} \frac{B_j}{3 \cdot 2^{j-1} - 1} \tag{1}$$

for all positive integers n, where B_j denotes the j^{th} Bernoulli number given by $B_0 = 1$ and $(j+1)B_j = -\sum_{m=0}^{j-1} \binom{j+1}{m} B_m$ for $j \geq 1$.

The Cantor set C is given by $[0, 1]\setminus \bigcup_{k=1}^{\infty} \bigcup_{j=1}^{2^{k-1}} I_{k,j}$, where $I_{1,1}$ is the open interval $(1/3, 2/3)$ and the open intervals $I_{k,1}, I_{k,2}, \ldots, I_{k,2^{k-1}}$ are the middle thirds of the 2^{k-1} component intervals of $[0, 1]\setminus \bigcup_{m=1}^{k-1} \bigcup_{j=1}^{2^{m-1}} I_{m,j}$. The $I_{k,j}$ are pairwise disjoint, F is constant on each $I_{k,j}$, and the range of F on $\bigcup_{j=1}^{2^{k-1}} I_{k,j}$ is given by $\{(2j-1)/2^k : 1 \leq j \leq 2^{k-1}\}$. Thus the function F takes the value 1/2 for $x \in [1/3, 2/3]$, an interval of length 1/3, the value 1/4 for $x \in [1/9, 2/9]$ and 3/4 for $x \in [7/9, 8/9]$, intervals of length 1/9, and so forth.

To prove (1), let $\sigma_n(m) = \sum_{j=1}^{m} j^n$. Note that

$$\sum_{j=1}^{2^{k-1}} (2j-1)^n = \sigma_n(2^k) - 2^n \sigma_n(2^{k-1}). \tag{2}$$

Since

$$\sigma_n(m) = \frac{1}{n+1} \sum_{j=0}^{n} \binom{n+1}{j} B_j (m+1)^{n+1-j}$$

(L. Comtet, *Advanced Combinatorics,* Riedel, 1974, p. 155), we have

$$\sigma_n(2^k) = 2^{kn} + \sigma_n(2^k - 1) = 2^{kn} + \frac{1}{n+1} \sum_{j=0}^{n} \binom{n+1}{j} B_j 2^{k(n+1-j)}. \tag{3}$$

The Cantor set C has measure zero, so we have

$$\int_0^1 (F(x))^n\,dx = \sum_{k=1}^{\infty} \sum_{j=1}^{2^{k-1}} \int_{I_{k,j}} (F(x))^n\,dx = \sum_{k=1}^{\infty} \frac{1}{3^k} \sum_{j=1}^{2^{k-1}} \left(\frac{2j-1}{2^k}\right)^n$$

$$= \sum_{k=1}^{\infty} \frac{\sigma_n(2^k) - 2^n \sigma_n(2^{k-1})}{3^k 2^{nk}} = \sum_{k=1}^{\infty} \frac{1}{3^k} \left(\frac{\sigma_n(2^k)}{2^{nk}} - \frac{\sigma_n(2^{k-1})}{2^{n(k-1)}}\right)$$

$$= \sum_{k=1}^{\infty} \frac{1}{3^k} \frac{\sigma_n(2^k)}{2^{nk}} - \frac{\sigma_n(1)}{3} - \sum_{k=1}^{\infty} \frac{1}{3^{k+1}} \frac{\sigma_n(2^k)}{2^{nk}} = -\frac{1}{3} + \frac{2}{3} \sum_{k=1}^{\infty} \frac{\sigma_n(2^k)}{3^k \cdot 2^{nk}}. \tag{4}$$

We have used (2) in going from the first line to the second. Substituting (3) into (4) yields

$$\begin{aligned}\int_0^1 (F(x))^n\,dx &= -\frac{1}{3}+\frac{2}{3}\sum_{k=1}^{\infty}\frac{1}{3^k}\left(1+\frac{1}{n+1}\sum_{j=0}^{n}\binom{n+1}{j}B_j 2^{(1-j)k}\right)\\ &= \frac{2}{3(n+1)}\sum_{k=1}^{\infty}\frac{1}{3^k}\sum_{j=0}^{n}\binom{n+1}{j}B_j 2^{(1-j)k},\end{aligned}$$

since $\sum_{k=1}^{\infty} 3^{-k} = 1/2$. An interchange in the order of summation now yields (1), since $\sum_{k=1}^{\infty}(3\cdot 2^{j-1})^{-k} = 1/(3\cdot 2^{j-1}-1)$. Putting $B_1 = -1/2$, $B_2 = 1/6$, $B_3 = 0$, and $B_4 = -1/30$ into (1) yields

$$\int_0^1 (F(x))^2\,dx = \frac{3}{10},\quad \int_0^1 (F(x))^3\,dx = \frac{1}{5},\quad \text{and}\quad \int_0^1 (F(x))^4\,dx = \frac{33}{230}.$$

Solution II by Omran Kouba, Higher Institute of Applied Sciences and Technology, Damascus, Syria. The function $F(x)$ satisfies the following self-similarity property: For every $x \in [0,1]$, we have

$$F(x) = 2F\left(\frac{x}{3}\right) = 2F\left(\frac{2}{3}+\frac{x}{3}\right)-1.$$

Let $A(t) = \int_0^1 \exp(tF(x))\,dx$ for $t \in \mathbb{R}$. Using the self-similarity property and $F(1/3) = F(2/3) = 1/2$ yields

$$\begin{aligned}A(2t) &= \int_0^{1/3}\exp(2tF(x))\,dx + \int_{1/3}^{2/3}\exp(2tF(x))\,dx + \int_{2/3}^{1}\exp(2tF(x))\,dx\\ &= \frac{1}{3}\int_0^1 \exp\left(2tF\left(\frac{x}{3}\right)\right)dx + \frac{1}{3}e^t + \frac{1}{3}\int_0^1 \exp\left(2tF\left(\frac{2}{3}+\frac{x}{3}\right)\right)dx\\ &= \frac{1}{3}(A(t)+e^t+e^tA(t)).\end{aligned}$$

Thus

$$1+3A(2t)-(1+e^t)(1+A(t)) = 0. \tag{5}$$

On the other hand, letting $J_n = \int_0^1 (F(x))^n\,dx$, we have $A(z) = \sum_{n=0}^{\infty} z^n J_n/n!$. Substituting this in (5) gives

$$\sum_{n=0}^{\infty}\left((3\cdot 2^n-1)J_n - 1 - \sum_{k=0}^{n}\binom{n}{k}J_k\right)\frac{t^n}{n!} = 0.$$

It follows that we may evaluate the sequence $(J_n)_{n\ge 0}$ by the recursion

$$J_0 = 1,\quad J_1 = \frac{1}{2},\quad \text{and}\quad J_n = \frac{1}{3\cdot 2^n - 2}\left(2+\sum_{k=1}^{n-1}\binom{n}{k}J_k\right)\quad \text{for}\quad n\ge 2.$$

Editorial comment. The recurrence is a special case of equation (5) of J. R. M. Hosking, Moments of order statistics of the Cantor distribution, *Stat. and Prob. Letters* **19** (1994) 161–165. Javier Duoandikoetxea notes that the integral $J_t = \int_0^1 (F(x))^t\,dx$ converges for all $t > -\log 3/\log 2$, and that $J_{-1} = \sum_{k=0}^{\infty} J_k$. Can the precise value of J_{-1} be computed?

Solved also by B. Burdick, R. J. Chapman (U. K.), J. E. Dawson (Australia), M. Desjarlais, J. Duoandikoetxea (Spain) T. Hermann, J. R. M. Hosking, J. H. Lindsey II, O. P. Lossers (Netherlands), V. Lucic (Canada), S. Mahajan, K. Schilling, N. C. Singer, A. Stenger, F. W. Steutel (Netherlands), D. C. Terr, A. Tissier (France), D. B. Tyler, Anchorage Math Solutions Group, GCHQ Problems Group (U. K.), WMC Problems Group, and the proposers.

REVIEWS

Edited by **Harold P. Boas**
Mathematics Department, Texas A & M University, College Station, TX 77843-3368

Notes on Fermat's Last Theorem. By Alf van der Poorten. Wiley-Interscience, 1996, xv + 222 pp., $49.95.

Reviewed by **Andrew Granville**

Have you ever wanted a math book that you could dip into like a favorite, inspired novel? One in which every page has a delicious quote, a provoking viewpoint, or a novel insight? A book that when read for the third time still makes you think or smile? A book that you can't put down, finding yourself reading on, even when you only picked it up to check on one little fact? This is Van der Poorten's polished, eccentric, opinionated, and inspiring *Notes on Fermat's Last Theorem.* We need more mathematics books like this.

Van der Poorten has written a book to inspire as many mathematicians as possible to enjoy the wonderful ideas behind modular forms and elliptic curves, and, in so doing, learn about much of mainstream number theory. He doesn't attempt to be complete, but instead tries to explain the flavour of much of what goes on:

> One of the difficulties in reading, or listening to, mature mathematics is its immense vocabulary and the volume of notions that seems to be required. Nor can one readily discover the meaning of the more popular ideas because all too often they are defined in terms of yet more obscure words. The truth is, unfortunately, that few—perhaps none—of us know all the definitions. We rely on a feeling for what must be intended, knowing that we can refine that feeling should needs be. In a sense, these notes should be seen precisely as an attempt to create some useful feelings.
>
> The style I have adopted in the notes is to *announce* all sorts of things. Some announcements are just definitions, others are facts whose explanations we are not yet in a position to comprehend. But many of my claims are indeed obvious after one has thought a little while....

Reading Van der Poorten is a bit like hearing a great colloquium in which you grasp the point of research in a field distant from your own, in part because the speaker astutely judges the correct amount of detail to present to persuade and interest you, and in part because the speaker assumes a level of rigour that allows you to follow and yet trust in what is going on, without being overwhelmed.

Here's an example from Lecture VII: Van der Poorten wishes to explain to the reader how we know that $\sin \pi z$ is periodic, given only that we have certain function theoretic properties of it, evidently so that he can later develop the techniques to understand elliptic functions, which are periodic in two directions on the plane. He writes

> ... we admit that $\sin \pi z$ has simple zeros exactly at $0, \pm 1, \pm 2, \ldots,$ and—rather wildly thinking of it as just a polynomial of infinite degree—we factorize it and write
>
> $$\sin \pi z = \pi z \prod_{n=1}^{\infty}\left(1 - \frac{z^2}{n^2}\right).$$

Of course that multiplier π (which, after all, might have been any decent function that never vanishes) needs rather calmer justification.

In a footnote, Van der Poorten provides this calmer justification by proving the existence of a set of points, dense on $[-1,1]$, that each satisfy this equation. He then proceeds:

With this evil deed done, we acknowledge that we are frightened of products, so we take the logarithm; and being bothered by logarithms, we differentiate. That yields

$$\pi \cot \pi z = \frac{1}{z} - \sum_{n=1}^{\infty} \left(\frac{1}{n-z} - \frac{1}{n+z} \right).$$

Unfortunately, as we catch our breath, we see that this is a mildly nasty partial fraction expansion in that it only converges conditionally—that is, on condition that we don't muck about with those parentheses. So we differentiate again and contemplate

$$\pi^2 \operatorname{cosec}^2 \pi z = \sum_{n=-\infty}^{\infty} \frac{1}{(n-z)^2}$$

and see that it shouts its periodicity. If we now backtrack carefully, we are done.

Much important technical mathematics is covered here, rigorously (though not pedantically!), but at the same time with an attempt to draw the reader's attention to the flow of the argument, rather than to distract the reader with the details. I find that this style draws me on, as a reader, inviting me to continue beyond what I know, on an easy trust, since I feel that the justification is there in a form that I can later revisit and try to understand.

To me this is in marked contrast to most mathematics books of today, which almost universally suffer from the Bourbakiist school of thought that everyone must speak the same highly technical language to appreciate what is going on (and so, those who do not know the jargon are doomed to not understand what is going on). In contrast, Van der Poorten's book demonstrates how one can get to the heart of the matter *without* dragging the reader through vast quantities of background material, presented in as dry a manner as possible. Why do so many authors act as if mathematics has to be such a very serious business, correct only if every "i" is dotted? That people might learn more, and more quickly, by being excited, inspired, and challenged to think about the questions of the day, has escaped this strangely dominant school of educational thought. However it hasn't escaped Van der Poorten's notice. His choice is to select topics that are fun, that give the flavour to some of the meat of the subject, and yet can be explained in a series of short, accessible chapters. Sometimes his explanations are not entirely rigorous, or need more justification later, but he candidly admits to these sins, and it makes me want to fill in the gaps, not to despair.

Most of the cognoscenti have shied away from writing books on the recent proof of Fermat's Last Theorem because of the difficulty in making such technical material accessible to the non-expert, while still doing justice to its profundity. A difficult task indeed, and all the other such books I have read fail dismally. However, Van der Poorten is perhaps the first such author to grasp even the basic material well enough to have the confidence to decide on a consistent, and plausible, perspective. What he does is to provide much of the basic background material, and the flavour of some of the less basic, without getting gum on his shoes, by being constructive, and sometimes by being intentionally redundant to

highlight an idea. Van der Poorten puts it well:

> I proceed to mention all sorts of odds and ends in an effort to sneak up on Wiles' argument without becoming too tangled in incomprehensible detail .... The point is to glimpse all sorts of exciting pieces of mathematics and to be moved to teach ourselves more. Among my motives in giving these lectures was that of trying to make mathematics a little less boring. All too often the reason for the incomprehensible things one is asked to learn is "beyond the scope of the text". That seems a constipated approach to me .... My idea was to provide motive—and damn the details.

It would be so beneficial if more authors dared to write in this way, but how can we re-direct our mathematics culture to make this less the extreme and more the norm?

All-in-all Van der Poorten's approach reminds me of when an excited colleague explains to me her latest research work in a field remote from mine, not being shy to discuss those little details that fascinate her (which are sometimes beside the point), while all the time throwing in lots of excellent examples, to make sure that I don't get totally lost.

Van der Poorten does not attempt to give a complete proof of Wiles' Theorem, nor to get involved with many of the difficult technical aspects. What he does do is to give a coherent overview of the proof, digging deep enough to include some essence from the harder mathematics involved, and to introduce various fundamental questions that arise. It is enough to get you started if you intend to go on to master the details.

The first few chapters of the book discuss much of the early history of number theory, in the guise of its relationship to Fermat's Last Theorem. Thus Van der Poorten covers much of the material central to Ribenboim's classic book **[1]**, but also gets to describe some of his own favourite topics, such as continued fractions and p-adic numbers. In chapter six he starts in on the modern approaches to Fermat's Last Theorem, introducing Mordell's and Faltings' Theorems, the *abc*-conjecture, and even a first shot at explaining the Birch–Swinnerton-Dyer conjectures. In chapters seven and ten, he introduces elliptic functions and Weierstrass parametrizations (including some nifty little tricks I'd not seen before), and some of the theory of Eisenstein series and modular forms. In the meantime in chapter nine he gives enough of the basics of reductions of elliptic curves that he can explain the modularity conjecture *accurately* in chapter eleven.

In chapter twelve he gives a lovely introduction to Poisson summation, which allows him to deduce functional equations and explain some of why they are interesting. In chapter thirteen he gives a more detailed discussion of L-functions and their role in modern mathematics. In chapter fifteen Van der Poorten discusses heights. This beautiful section gives clear motivation to view several notions of height as aspects of the same idea, proceeding from Mahler measure to canonical height. This then allows him to give a more complete explanation of the Birch–Swinnerton-Dyer conjectures in chapter sixteen. By describing the construction of Heegner points, Van der Poorten then explains the Gross-Zagier formula, and so motivates the solution to Gauss's class number problem. In chapter seventeen, Van der Poorten has a stab at explaining the relevance of Galois representations, the Deligne-Serre theorem, and thence goes on to his sketch of the proof of Wiles. Once you have gotten this far, you are ready to move on to learning more of the details, and Van der Poorten has succeeded in his goal of getting you *involved* in the mathematics.

One beautiful aspect of this book is that Van der Poorten manages to discuss several interesting recent developments in number theory that stand apart from Fermat's Last Theorem and the modularity conjecture. By developing number theory as he does, he shows how these questions arise naturally in their own context, and how the tools he discusses provide approaches to them.

Van der Poorten's book will be a special addition to your bookshelf not only for the discussion of mathematics, and for the refreshingly honest approach to how mathematics is really done, but also because he knows and discusses the people involved, he has a varied sense of humour and a freeish style of writing, and he is very aware of the cultural impact of the resolution of Fermat's Last Theorem.

Indeed, now that Fermat's Last Theorem has been proved (by the way, "Last" is as in "Last to be proved"), many people, including the MONTHLY book reviews editor Underwood Dudley who commissioned this article, are asking "What next?" What question is going to take the place of Fermat's Last Theorem, to inspire and provoke the next hundred generations of students? To inspire and provoke them to explore mathematics for themselves, to experiment, to play, and to discover that the more one probes in mathematics, the more one learns that there is so much yet to be understood? There are several candidates for such a question: old favourites such as the *Riemann Hypothesis*, the *Twin Prime Conjecture*, or the *Poincaré Conjecture*; more modern questions like $\mathbf{P} \neq \mathbf{NP}$ or the *abc-conjecture*; or off-shoots of Wiles' Theorem, such as *Prove that there are no coprime positive integers* x, y, z *satisfying* $x^p + y^q = z^r$ *with* $p, q, r \geq 3$. The experts, by definition, are unlikely to predict which question will turn out to inspire the uninitiated, and how it will provoke the next Wiles into becoming a mathematical researcher. In fact I doubt that any such question will emerge in the foreseeable future.

Having been asked this question repeatedly in the last few years, I have tried to explore the impact of Wiles' extraordinary work on the culture that defines our subject. I have come to the conclusion that the proof of Fermat's Last Theorem is as much Mathematics' greatest loss as one of Mathematics' greatest wins. Fermat's Last Theorem had everything: a romantic story (the lost, and marginal, proof), easily understood background information (Pythagoras' Theorem, and 3-4-5 triangles), a misleading appearance of accessibility (all those flawed proofs), a deep and rich mathematical history (Cauchy, Kummer, Vandiver, ..., Faltings, Frey, Serre, Ribet, ...), early work from one of the greatest of the early female mathematicians (Germain), and even high financial stakes (the Wolfskehl prize)!

Now that the Holy Grail of Mathematics has been found, how else to rally the faithful to Camelot? Can the deeper, more complex, and arguably more important questions entice? Surely the uninitiated always have wanted and always will want a glittering prize ahead of them, just out of reach, but close enough to draw them on? Perhaps it would have been better if Fermat's Last Theorem had never been resolved, if it had remained as testament to our limitations, just beyond the reach of mortal ken.

It has long been trendy to downplay the importance of Fermat's Last Theorem, and instead to focus on deeper, less immediately enticing questions. Indeed Kummer called Fermat's Last Theorem "more of a joke than a pinnacle of science", while Gauss would not deign to work on it (or so he claimed!). However, I believe that behind this sophisticated façade, either would have been delighted to resolve the question (clearly much of Kummer's greatest work was motivated by his study of Fermat's Last Theorem). Several of today's expert naysayers, who ridiculed the significance of Fermat's Last Theorem not so long ago, put aside their prejudices when the time came, and enthusiastically rejoiced in the amazing

conclusion to the Fermat story. This was fitting, for Wiles' great proof is a true milestone in the history of mathematics: it exhibits how so many of the abstract developments of mathematics influence the simplest of all serious questions, and perhaps let us believe that so much of the work that has been done, in so many diverse areas, is really worthwhile!

How can any other question reflect so well the mathematical culture from which it springs, and in which it is finally laid to rest?

Finally, let me repeat that Van der Poorten's monograph is a wonderful mathematics book, which dares to breach the stylistic barriers that usually impede understanding. It encompasses a lot of material, from most elementary to very deep, but remains accessible. I expect it will turn a lot of people on to number theory and arithmetic geometry, and indeed the beauty of mathematics as a whole. At the very least, if you have a clever undergraduate student who is bored by upper division calculus and ready for something a little more poignant, get her to read this book, and let her first experience of research-level mathematics be provoking, inspiring, and fun.

REFERENCES

1. Paulo Ribenboim, *13 Lectures on Fermat's Last Theorem*, Springer-Verlag, New York, 1979.
2. Andrew Wiles, Modular elliptic curves and Fermat's last theorem, *Annals of Math.* (2) **141** (1995) 443–551.

University of Georgia, Athens, GA 30602-7403
andrew@math.uga.edu

The Basic Practice of Statistics. By David S. Moore. W. H. Freeman and Company, New York, 1995, 680 pp., $59.95.

Reviewed by Frederick Leysieffer

It finally happened. Your colleague who has been teaching the statistics courses all these years is going on leave. Your chair looked at the transcripts of everyone in your department and discovered that you once took a course in statistics. Now you are responsible for the departmental statistics teaching next year. Your first step is to select a text. What do you look for?

The first thing to remember is that statistics is a discipline in and of itself. Sure, lots of mathematics departments teach statistics, but so do departments of psychology, education, economics, industrial engineering, and so on. And, of course, so do departments of statistics. Statistics as a discipline derives its reason for existence and its methodological impetus from measurement problems in the many and varied areas of application, and it relies on rigorous mathematics for its theoretical foundation.

To one accustomed to the beauties of mathematical rigor, books for a first course in statistics appear to have too much of a cookbook flavor. Often they follow the lines of: "Here is a data set. Analyze it this way." No reasons are given. All statistics books that seek to describe various methods of data analysis to

non-mathematical students have to invoke this approach to some extent. To do otherwise would require students to have a more extensive mathematical background in order to understand the derivations of formulas. This approach is consistent with the background that many users of statistics have. In fact, many very effective users of statistical techniques take the mathematical derivations on faith along with the statement of assumptions that must be satisfied before a procedure can be used validly. They then apply statistical techniques to problems in the social and natural sciences.

There is, however, a type of rigor one absolutely must have in elementary texts. One must demand clarity and precision in the presentation of definitions of statistical concepts. Notions such as the laws of large numbers, the central limit theorem, sampling distributions, p-values, and the meaning of confidence intervals are examples of concepts that require careful definition to facilitate clear thinking and comprehension by students. These concepts all have precise meanings harking back to their mathematical origins. They are not easy for students to grasp. Muddy definitions are useless. Carefully crafted qualitative definitions are necessary and possible in a text at this level.

Furthermore, as each statistical procedure is introduced, the assumptions that one must make about the data to validate the use of the procedure must be clearly and unambiguously stated. The exposition of these points must not be muddy.

Teaching statistics should be and certainly can be fun. I prefer texts from which I can learn, texts in which the author's examples come from actual situations. Examples or exercises that begin: "A certain manufacturer produces widgets..." are dull and really deadly if you want to keep the interest of your students. Avoid texts with cutesy examples and go for the texts with real data. There are many fascinating data sets out there—after all, statistics is about data analysis. Don't settle for the ersatz. Make sure there are lots of examples and lots of exercises to work.

The order of presentation of material makes a difference in orienting your students, and indeed in how you can teach your course. The standard formula, used in many well-regarded texts, is to begin with a review of set theory, to introduce elementary probability theory, and gradually to work up to some inference problems, usually testing hypotheses and estimation through confidence intervals.

The problem with that progression of topics for the less mathematically prepared student is that although the order is mathematically logical, you really do not get around to studying statistics until it is too late. There are simply not that many interesting real-life examples illustrating the addition rule for probabilities. Students who thought they signed up to learn about data analysis lose interest. The old standard ordering of material does not really drive home the point that statistics is a separate discipline. It provides a bit of mathematical rigor, but only through some really elementary mathematics, after which it reverts to the cookbook style again.

A better way is to begin with statistical ideas. There is much one can do to give students a feeling for statistics before introducing probabilities and doing formal inference. One can consider data displays and examine relationships between variables without probabilities. It makes sense to consider sampling and the collection of data as a valid component of a statistics course. With engaging examples, students stay interested.

These are all reasons why *The Basic Practice of Statistics* by David S. Moore is well worth considering if your course introduces statistical ideas to students having

minimal mathematical preparation. Moore was president of the American Statistical Association in 1998, and he has served on a joint committee of that organization and the Mathematical Association of America that studied the teaching of introductory statistics. His objective, as stated in the book's introduction, is to emphasize statistical thinking with more data, more concepts, less theory, and fewer recipes. His recent MONTHLY paper with Cobb [**2**] gives a comprehensive exposition of this approach. As evidence of his dedication to this methodology, the first three chapters of the text—225 pages—assume no probability theory.

The material on graphical displays in the first chapter is the tip of the iceberg of exploratory data analysis. The second chapter gives examples of the difficult problem of finding relationships in large data sets of the sort that scientists (meteorologists, for example) routinely generate in their research. Effective visual displays can help in examining data sets: they reveal relationships that otherwise may be obscured by the wealth of data collected. For ancillary reading, students can refer to Tufte [**8**] for a beautiful exposition of visual displays of data.

The chapter on sampling is welcome in an introductory text. Careful data collection is an art. It is full of pitfalls, some predictable and others unexpected. Students become better consumers of statistics if they realize this and understand basic criteria for effective data collection, notions of randomness, principles of experimental design, and ethical considerations in the collection process—goals that one might hope to achieve in an introductory course in statistics.

Group projects provide an excellent vehicle to introduce students to the processes of effective and useful data collection. There is a difference between reading about data collection and actually doing it. Projects give a statistics course the added dimension of practical experience. The ordering of topics in this text lends itself extremely well to doing class projects. Early on, students who have mastered the first three chapters can decide on a project, determine what they want to measure, design a data collection procedure, display their results, and begin to look for relationships in the data. With careful coaching by the instructor, students can collect data that can then be analyzed with the inference techniques to be learned later in the course.

In the fourth chapter, Moore introduces sampling distributions in an intuitive way, motivating the subsequent discussion of probability distributions by appealing to empirical considerations. This is a key chapter since all that follows depends upon one's accepting, at this level, the various sampling distributions that go with the variety of data analysis procedures. We do not see set theory, nor is there an emphasis on rules for probabilities.

With the fifth chapter and beyond, the text settles into a more formula-based mode that requires students to accept the validity of the various sampling distributions. At this point, mathematically prepared students who know about multiple integrals and change of variables and who are curious about the origins of the t- or F-distributions or similar concepts could consult a mathematical statistics text for derivations. The classic text by Hogg and Craig [**4**] is a good place to start.

Subject coverage is certainly appropriate for a text at this level. The essential concepts of statistical inference are treated, though nonparametric statistics are not covered.

The instructor may want to supplement the course by showing parts of the video series *Against All Odds* [**1**]. Moore was the content developer for the series, and it meshes nicely with the material in the text. See Moore's article on the place of video materials in the classroom [**5**]. Also, a recently issued CD-ROM and workbook, *An Electronic Companion to Statistics* [**3**], features short parts of the series.

In considering this text, one could also look at two other books by Moore that bracket this one in the amount of mathematical preparation required. *Statistics: Concepts and Controversies* [6] takes a more qualitative approach and emphasizes statistical concepts over techniques. Moore's book with McCabe, *Introduction to the Practice of Statistics* [7], emphasizes a bit more of the formal development and is intended for a more advanced student. Either book could be helpful ancillary reading for students.

Numerous fine examples throughout the text keep up the interest of both the instructor and the student. Some examples can surprise the reader, and some can reinforce previous beliefs. A scatter plot shows the inverse relationship between the median SAT score in a state and the percentage of high school seniors in the state who took the test (p. 103). There is an example of a lurking variable related to public housing in Hull, England (p. 144). The notes and data sources sections at the ends of chapters exhibit the wide variety of disciplines from which real data were chosen for illustrative purposes.

The text is pleasantly laid out. The exposition is excellent and precise. Moore includes nice ancillary features in his text. Warnings illustrate when statistical inference is not valid for all sets of data, e.g., the discussion of the Hawthorne effect on p. 383. The biographical sketches of famous statisticians inject welcome human interest.

However one structures a first course in statistics, if the student audience is minimally prepared mathematically, then David Moore's carefully crafted text deserves attention.

REFERENCES

1. *Against All Odds: Inside Statistics*, series of 26 half-hour videos on 13 cassettes, produced by the Consortium for Mathematics and Its Applications and the Chedd-Angier Production Company, 1989, distributed by The Annenberg/Corporation for Public Broadcasting Project, S. Burlington, Vermont.
2. George W. Cobb and David S. Moore, Mathematics, statistics, and teaching, *Amer. Math. Monthly* **104** (1997) 801–823.
3. George W. Cobb, Jeffrey A. Witmer, and Jonathan D. Cryer, *An Electronic Companion to Statistics*, Cogito Learning Media, 1997.
4. Robert V. Hogg and Allen T. Craig, *Introduction to Mathematical Statistics*, fifth edition, Prentice Hall, 1995.
5. David S. Moore, The place of video in new styles of teaching and learning statistics, *Amer. Statist.* **47** (1993) 172–176.
6. David S. Moore, *Statistics: Concepts and Controversies*, fourth edition, Freeman, 1997.
7. David S. Moore and George P. McCabe, *Introduction to the Practice of Statistics*, third edition, Freeman, 1998.
8. Edward R. Tufte, *The Visual Display of Quantitative Information*, Graphics Press, 1983, reprinted 1997.

Office of the Provost, The Florida State University, Tallahassee, FL 32306
leysieff@stat.fsu.edu

TELEGRAPHIC REVIEWS

Edited by **Arnold Ostebee**
with the assistance of the Mathematics Departments of
Carleton, Macalester, and St. Olaf Colleges

Telegraphic Reviews are designed to alert readers in a timely manner to new books appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T : Textbook	*P : Professional Reading*	*1–4: Semester*
C : Computer Software	*L : Undergraduate Library*	*** : Special Emphasis*
S : Supplementary Reading	*13: Grade Level*	*?? : Questionable*

Readers are advised that price information is subject to change. Selected books receive a second, more extensive review in the *Monthly.*

Books submitted for review should be sent to *Book Reviews Editor, American Mathematical Monthly, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098.*

General, P, L*.** *Handbook of Writing for the Mathematical Sciences, Second Edition.* Nicholas J. Higham. SIAM, 1998, xvi + 302 pp, $34 (P). [ISBN 0-89871-420-6] Extensively revised and updated; several new chapters and some new sections. Well-organized, excellent advice enlivened with humor. Every mathematician should have a copy within easy reach. (*First Edition*, TR, April 1994.) AO

General, P*, L*. *The World According to Wavelets: The Story of a Mathematical Technique in the Making, Second Edition.* Barbara Burke Hubbard. AK Peters, 1998, xx + 330 pp, $40. [ISBN 1-56881-072-5] New material on the history and applications of wavelets; updated bibliography and list of web sites. Well-written. Accessible to undergraduates. (*First Edition*, TR, January 1997.) AO

Reference, S(14–16), P, L. *Mathematical Methods for Physics and Engineering.* K.F. Riley, M.P. Hobson, S.J. Bence. Cambridge Univ Pr, 1997, xix + 1008 pp, $49.95 (P); $110. [ISBN 0-521-55529-9; 0-521-55506-X] Brief expositions of a wide variety of mathematical tools including basic calculus, ODEs, PDEs, complex analysis, tensors, calculus of variations, integral equations, group theory, probability, numerical methods, and more. Each topic is introduced qualitatively, then more rigorously. Includes numerous worked examples, exercises, hints, and answers. AO

Precalculus, T(13: 2), S*. *Precalculus in Context: Projects for the Real World, Second Edition.* Marsha J. Davis, Judith Flagg Moran, Mary E. Murphy. Brooks/Cole, 1998, xvii + 311 pp, $25.95 (P). [ISBN 0-534-35232-4] Thirteen labs designed for group collaboration and exploration of typical precalculus topics; each lab is followed by more guided explorations and projects. Uses graphing technology extensively. MW

Education, P, L*.** *Writing in the Teaching and Learning of Mathematics.* John Meier, Thomas Rishel. MAA Notes No. 48. MAA, 1998, xiii + 100 pp, $18.95 (P). [ISBN 0-88385-158-X] Reflections on why to use writing assignments as well as advice on creating effective assignments. Most sections include exercises/discussion questions. Easy to read; numerous examples illustrate approaches and pitfalls. An important and valuable resource. AO

Education, P. *Teacher-Made Aids for Elementary School Mathematics, Volume 3.* Ed: Carole J. Reesink. NCTM, 1998, vi + 377 pp, $17.95 (P). [ISBN 0-87353-463-8] A collection of articles published in the *Arithmetic Teacher* and *Teaching Children Mathematics* between 1984 and 1997.

Education, P*, L.** *History of Mathematics: Histories of Problems.* The Inter-IREM Commission. Transl: Chris Weeks. Ellipses (Edition Marketing S.A., 32 rue Bargue, 75740 Paris, Cedex 15), 1997, 429 pp, 220F. [ISBN 2-7298-4730-8] A "great problems" approach to the history and culture of mathematics aimed at teachers. Each chapter explores the historical evolution of a single problem and the tools developed for its solution. Includes historical material, bibliographies, exercises. AO

Education, S, L. *Twenty Years Before the*

Blackboard: The Lessons and Humor of a Mathematics Teacher. Michael Stueben with Diane Sandford. Spectrum Ser. MAA, 1998, xi + 155 pp, $29.50 (P). [ISBN 0-88385-525-9] A potpourri of stories, maxims, fables, trifles, paradoxes, proofs, and jests drawn from the author's two decades of experience as a high school mathematics teacher. A mildly useful source of enrichment for teachers of today. LAS

Education, P. *Perspectives on the Teaching of Geometry for the 21st Century: An ICMI Study.* Eds: Carmelo Mammana, Vinicio Villani. ICMI Study Ser., V. 5. Kluwer Academic, 1998, viii + 353 pp, $129. [ISBN 0-7923-4990-3] Outgrowth of ICME–7 (1992) and subsequent Catania Conference (1995). Highlights major trends and improvements in geometry teaching. Traces evolution of geometry curricula in several countries. Addresses issues in assessment and teacher education. MW

Education, P, L. *Language and Communication in the Mathematics Classroom.* Eds: Heinz Steinbring, Maria G. Bartolini Bussi, Anna Sierpinska. NCTM, 1998, vii + 351 pp, $11.95 (P). [ISBN 0-87353-441-7] Twenty papers from teachers and researchers around the world derived from presentations at ICME–7 (1992) in Quebec. Stresses the cultural roots of communication and their implications for discussion in mathematics classes. "The meanings of mathematical terms ... are not merely linguistic: they are cultural." LAS

History, P. *Collected Papers of Norman Alling.* Ed: Paulo Ribenboim. Papers in Pure & Appl. Math., V. 107. Queen's Univ, 1998, xii + 597 pp, (P). [ISBN 088941-796-9]

History, P, L*. *Trigonometric Delights.* Eli Maor. Princeton Univ Pr, 1998, xiv + 236 pp, $24.95. [ISBN 0-691-05754-0] An enjoyable blend of mathematics, history, and biography. Most of the book is accessible to readers with only a basic background in high school algebra and trigonometry; the remainder assumes some knowledge of first-year calculus. AO

Linear Algebra, T*(13–14: 1). *The Keys to Linear Algebra: Applications, Theory, and Reasoning.* Daniel Solow. BookMasters (PO Box 388, 1444 St. Route 42, Ashland, OH 44805), 1998, ix + 548 pp, $49.95. [ISBN 0-9644519-2-1] Aims to teach the thinking process. Each chapter is organized around a motivating application. Exercises on mechanics, theory, and technology-based problem solving. TH

Algebra, T(18: 2), P, L. *Representations and Invariants of the Classical Groups.* Roe Goodman, Nolan R. Wallach. Ency. of Math. & Its Applic., V. 68. Cambridge Univ Pr, 1998, xvi + 685 pp, $100. [ISBN 0-521-58273-3] Updates and continues the theory found in Weyl's *The Classical Groups* (TR, April 1998). Abundant examples and exercises. TH

Algebra, P. *The Book of Involutions.* Max-Albert Knus, *et al.* Colloq. Pub., V. 44. AMS, 1998, xxi + 593 pp, $69. [ISBN 0-8218-0904-0] Theory of central simple algebras with involution in relation with linear algebraic groups.

Algebra, P. *Abelian Groups, Module Theory, and Topology: Proceedings in Honour of Adalberto Orsatti's 60th Birthday.* Eds: Dikran Dikranjan, Luigi Salce. Lect. Notes in Pure & Appl. Math., V. 201. Marcel Dekker, 1998, xv + 444 pp, $165 (P). [ISBN 0-8247-1937-9] Papers from a 1997 conference in Padua, Italy.

Algebra, T(15–17: 1), L*. *Discrete Mathematics Using Latin Squares.* Charles F. Laywine, Gary L. Mullen. Ser. in Disc. Math. & Optimiz. Wiley, 1998, xv + 305 pp, $79.95. [ISBN 0-471-24064-8] Theory of mutually orthogonal Latin squares with applications to combinatorics, statistics, error-correcting codes, and cryptology. Assumes some familiarity with finite fields. DB

Calculus, T(13: 2), L. *Modeling the Dynamics of Life: Calculus and Probability for Life Scientists.* Frederick R. Adler. Brooks/Cole, 1998, xvi + 784 pp, $73.95. [ISBN 0-534-34816-5] Written to introduce life science majors to concepts fundamental to theoretical biology. Uses themes of growth, diffusion, and selection to present three types of models: discrete-time dynamical systems, differential equations, stochastic processes. Covers many traditional calculus topics as well as some from probability and statistics. AO

Calculus, S*(14), L*. *Div, Grad, Curl, and All That: An Informal Text on Vector Calculus, Third Edition.* H.M. Schey. WW Norton, 1997, ix + 164 pp, $20 (P). [ISBN 0-393-96997-5] Minor textual changes, corrections, and several new exercises. Major change is replacement of the operators div, grad, and curl by $\nabla\cdot$, ∇, and $\nabla\times$, respectively. (*First Edition*, TR, November 1973.) AO

Differential Equations, S(14). *Solving Differential Equations with MAPLE V Release 4.* David Barrow, *et al.* Brooks/Cole, 1998, viii + 253 pp, $21.50 (P). [ISBN 0-534-34555-7] A guide to using Maple's symbolic, graphical, and numerical capabilities. Introduces basic commands for dealing with differential equations, illustrates their use in applied contexts. Final chapter contains several projects. AO

Differential Equations, T(16–17). *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations.* Uri M. Ascher, Linda R. Petzold. SIAM, 1998, xvii + 314 pp, $36.50 (P). [ISBN 0-89871-412-5] Text for a first course on the numerical solution of ODEs. Covers both initial-value and boundary-value problems as well as differential-algebraic equations. A practical introduction to basic methods and issues. AO

Differential Equations, T(14: 1). *Introduction to Ordinary Differential Equations.* Stephen H. Saperstone. Brooks/Cole, 1998, xiv + 634 pp, $69.95. [ISBN 0-314-05819-2] Uses visualization tools, numerical estimation, and symbolic computation (including a computer algebra system). Covers all the traditional topics with somewhat more than the usual emphasis on modeling. AO

Differential Equations, T(14: 1).** *Differential Equations.* Paul Blanchard, Robert L. Devaney, Glen R. Hall. Brooks/Cole, 1998, xv + 732 pp, $62.95. [ISBN 0-534-34550-6] A substantial departure from traditional approaches to the subject; based on a dynamical systems point of view. Uses qualitative, numerical, and analytic approaches throughout. Worth a thoughtful look! AO

Differential Equations, T*(14: 1). *Differential Equations: A Modeling Perspective.* Robert L. Borrelli, Courtney S. Coleman. Wiley, 1998, xiii + 706 pp, $86.95. [ISBN 0-471-04230-7] Covers basic theory and closed form solution techniques as well as topics motivated by the modern view of differential systems as evolving dynamical systems (e.g., long-term behavior, sensitivity, bifurcation). Modeling and graphical visualization are central themes. Assumes access to a numerical solver. AO

Dynamical Systems, S(18). *Dynamical Systems and Ergodic Theory.* Mark Pollicott, Michiko Yuri. London Math. Soc. Stud. Texts, V. 40. Cambridge Univ Pr, 1998, xiii + 179 pp, $22.95 (P); $59.95. [ISBN 0-521-57599-0; 0-521-57294-0] An introduction to topological dynamics and ergodic theory. Useful as a reference for those who already have a firm understanding of the basic topics and their overall organization. Very few applications. DK

Dynamical Systems, T(16–17: 2), P. *Elements of Applied Bifurcation Theory, Second Edition.* Yuri A. Kuznetsov. Appl. Math.. Sci., V. 112. Springer-Verlag, 1998, xix + 591 pp, $69.95. [ISBN 0-387-98382-1] An introduction to nonlinear dynamical systems. This edition contains new and improved numerical methods for bifurcation analysis, more complete analysis of codim 1 bifurcations (*First Edition*, TR, March 1996). SN

Numerical Analysis, S*(17), P, L*. *Afternotes Goes to Graduate School: Lectures on Advanced Numerical Analysis.* G.W. Stewart. SIAM, 1998, xii + 245 pp, $35 (P). [ISBN 0-87871-404-4] Topics: discrete and continuous approximation, linear and quadratic splines, eigensystems, and Krylov sequence methods. Also notes from two lectures on classical iterative methods and nonlinear equations. AO

Differential Geometry, P. *Geometry of Differential Equations.* Eds: A. Khovanskiĭ, A. Varchenko, V. Vassiliev. AMS Transl. Ser. 2, V. 186. AMS, 1998, xi + 194 pp, $89. [ISBN 0-8218-1094-4] 7 articles written by colleagues of V.I. Arnold in honor of his 60th birthday.

Geometry, T(15–16: 2). *Modern Geometries, Fifth Edition.* James R. Smart. Brooks/Cole, 1998, xvii + 451 pp, $68.95. [ISBN 0-534-35188-3] Fine tuning includes increase in number of exercises, expanded sections on fractals, updates to bibliography, references to dynamic geometry software (*Fourth Edition*, TR, August–September 1994). MW

Geometry, T(1), S*(15–16). *Active Geometry.* David A. Thomas. Brooks/Cole, 1998, 139 pp, $20.75 (P). [ISBN 0-543-34485-2] Hands-on investigations in both Euclidean and non-Euclidean geometries using computer modeling and analysis tools. Especially useful in courses for teachers. MW

Geometry, T(14), L. *The Geometry Toolbox for Graphics and Modeling.* Gerald E. Farin, Dianne Hansford. AK Peters, 1998, xv + 288 pp, $48. [ISBN 1-56881-074-1] Highly visual approach to basic two- and three-dimensional geometry and linear algebra. Approaches linear algebra from a geometric viewpoint, and geometry from an algorithmic viewpoint. Useful background for courses in computer graphics and geometric modeling. AO

Topology, S(16–18), L. *The Four-Color Theorem: History, Topological Foundations, and Idea of Proof.* Rudolf Fritsch, Gerda Fritsch. Transl: Julie Peschke. Springer-Verlag, 1998, xvi + 260 pp, $29.95. [ISBN 0-387-98497-6] A thorough retrospective on the four-color drama including biographies of key actors (Guthrie, deMorgan, Hamilton, ... Heesch, Haken, Appel), careful topological foundations (e.g., on arcwise accessibility) to rule out fractal "monsters," and a thorough exposition of the proof (reducibility, obstructions, discharging procedures). LAS

Optimization, T(17: 1), P, L. *Interior Point*

Algorithms: Theory and Analysis. Yinyu Ye. Ser. in Disc. Math. & Optim. Wiley, 1997, xvi + 418 pp, $74.95. [ISBN 0-471-17420-3] Exposition of the theory, analysis, and implementation of interior point methods. Emphasizes geometry and potential theory for convex inequalities. AO

Optimization, P. *New Trends in Mathematical Programming: Homage to Steven Vajda.* Eds: Franco Giannessi, Sándor Komlósi, Tamás Rapcsák. Appl. Optim., V. 13. Kluwer Academic, 1998, xi + 314 pp, $140. [ISBN 0-7923-5036-7] 22 papers on recent advances.

Optimization, T(18), P*, L. *Theory and Algorithms for Linear Optimization: An Interior Point Approach.* C. Roos, T. Terlaky, J.-Ph. Vial. Ser. in Disc. Math. & Optim. Wiley, 1997, xxiv + 482 pp, $100. [ISBN 0-471-95676-7] An interior point approach to both theoretical and algorithmic aspects of linear optimization. First part presents duality theory and a polynomial method for solving linear optimization problems. Remainder discusses logarithmic barrier methods, target-following approaches, and various other topics of theoretical and practical interest. AO

Optimization, P. *Minimax Theory and Applications.* Eds: Biagio Ricceri, Stephen Simons. Nonconvex Optim. & Its Applic., V. 26. Kluwer Academic, 1998, xi + 271 pp, $119. [ISBN 0-7923-5064-2] Proceedings of a 1996 workshop held in Erice, Italy.

Game Theory, S*(16–18). *Evolutionary Games and Population Dynamics.* Josef Hofbauer, Karl Sigmund. Cambridge Univ Pr, 1998, xxvii + 323 pp, $27.95 (P); $69.95. [ISBN 0-521-62570-X; 0-521-62365-0] A fascinating combination of dynamical systems and game theory in the service of understanding evolution and population dynamics. Suitable for either a mathematician interested in learning about biology or a mathematically knowledgeable biologist interested in learning about theoretical approaches. DK

Mathematical Modeling, T(13–14: 1), L. *A First Course in Mathematical Modeling, Second Edition.* Frank R. Giordano, Maurice D. Weir, William P. Fox. Brooks/Cole, 1997, xviii + 525 pp, $62.50. [ISBN 0-534-22248-X] This edition adds chapters on discrete dynamical systems, linear programming and numerical search methods, and an introduction to probabilistic modeling. Emphasizes model construction process. (*First Edition*, TR, August–September 1985.) AO

Mathematical Computing, P. *Essential MATLAB for Scientists and Engineers.* Brian D. Hahn. Wiley, 1997, x + 265 pp, $39.95 (P). [ISBN 0-470-25013-5] Offers a "teach-yourself" approach to using *MATLAB.* Does not assume prior knowledge of matrices or programming techniques.

Computer Science, P. *Logic Programming.* Ed: Joxan Jaffar. MIT Pr, 1998, xvii + 361 pp, $85 (P). [ISBN 0-262-60031-5] Proceedings of the 1998 Joint International Conference and Symposium on Logic Programming held at the University of Manchester, U.K.

Applications (Physics), T(17: 2). *Classical Dynamics: A Contemporary Approach.* Jorge V. José, Eugene J. Saletan. Cambridge Univ Pr, 1998, xxv + 670 pp, $54.95 (P); $125. [ISBN 0-521-63636-1; 0-521-63176-9] Comprehensive treatment of dynamics from a modern perspective. Solid treatments of Lagrangian and Hamiltonian dynamics, scattering and linear oscillators, canonical transforms, Liouville and Darboux theorems, non-linear dynamics (chaos, KAM theorem), continuum dynamics (chaos, solitons, Navier–Stokes), etc. Many interesting worked-out examples. RM

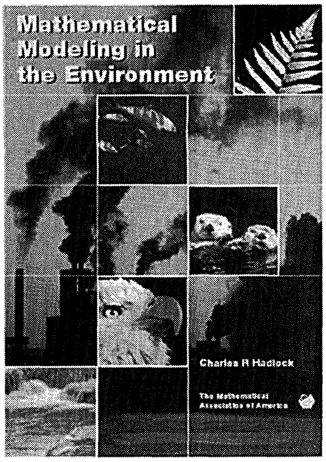
Applications (Quantum Theory), P. *Secondary Calculus and Cohomological Physics.* Eds: Marc Henneaux, Joseph Krasil'shchik, Alexandre Vinogradov. Contemp. Math., V. 219. AMS, 1998, xiv + 287 pp, $60 (P). [ISBN 0-8218-0828-1] Proceedings of a 1997 conference held at Moscow State University (Russia).

Applications (Quantum Theory), P. *Coherent Transform, Quantization, and Poisson Geometry.* Ed: M.V. Karasev. AMS Transl. Ser. 2, V. 187. AMS, 1998, x + 360 pp, $119. [ISBN 0-8218-1178-9] Three articles on problems in quantization theory, as well as in symplectic and Poisson geometry.

Applications, P. *Lecture Notes in Control and Information Sciences–236: Autonomous Robotic Systems.* Eds: Anibal T. de Almeida, Oussama Khatib. Springer-Verlag, 1998, viii + 275 pp, $79 (P). [ISBN 1-85233-036-8] Tutorial papers from a 1997 workshop at the University of Coimbra (Portugal). In four sections: Sensors and Navigation; Cooperation and Telerobotics; Applications; Legged and Climbing Robots.

Reviewers

DB: David Bressoud, Macalester; TH: Tom Halverson, Macalester; DK: Danny Kaplan, Macalester; RM: Richard Molnar, Macalester; SN: Sam Northshield, Carleton; AO: Arnold Ostebee, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MW: Martha Wallace, St. Olaf.



THE MATHEMATICAL ASSOCIATION OF AMERICA

Mathematical Modeling in the Environment

Charles Hadlock

Series: Classroom Resource Materials

Packaged with a PC compatible disk that enhances the material in the text.

Suitable for classroom adoption in an innovative course for

- **a general education mathematics elective**
- **a mathematics or science major advanced elective**
- **an interdisciplinary course, even at a relatively elementary level**
- **a mathematical modeling course in a civil/environmental engineering program**

This book has a dual objective: first, to introduce the reader to some of the most important and widespread environmental issues of the day; and second, to illustrate the vital role played by mathematical models in investigating these issues. The environmental issues addressed include: ground-water contamination, air pollution, and hazardous material emergencies. These issues are presented in their full real-world context, not as scientific or mathematical abstractions; and for background, readers are invited to investigate their status in their own communities.

The first part of the book leads the reader through relatively elementary modeling of these phenomena, including simple algebraic equations for ground water, slightly more complex algebraic equations (preferably implemented on a spreadsheet or other computerized framework) for air pollution, and a fully computerized modeling package for hazardous materials incident analysis. The interplay between physical intuition and mathematical analysis is emphasized.

For more advanced readers, the second part of the book returns to the same three subjects but with a higher level of mathematical sophistication (adjustable to the preparation of the reader by selection of subsections.) Many important classical mathematical themes are developed through this context, examples coming from single and multivariable calculus, differential equations, numerical analysis, linear algebra and probability. The material is presented in such a way as to minimize the required background and to encourage the subsequent study of some of these fields.

An elementary course for a general audience could be based entirely on Part I, and a higher level mathematics, science, or engineering course could move quickly to Part 2.

A PC compatible diskette packaged with the text contains a spreadsheet program that facilitates the numerical experimentation with the Gaussian plume equation introduced in Chapter 3, as well as public domain DOS program (ARCHIE) for evaluating the consequences from various hazardous materials scenarios (e.g., the physical extent of flammable and toxic vapor clouds). Text is not tied to the use of this software, but it is included as an aid to meet the pedagogical objectives of the text.

Catalog Code: ENV/SA
312 pp., Paperbound, 1998, ISBN 0-88385-709-X
List: $55.00 MAA Member: $43.95

Instructor's and Solutions Manual for *Mathematical Modeling in the Environment*

Charles Hadlock

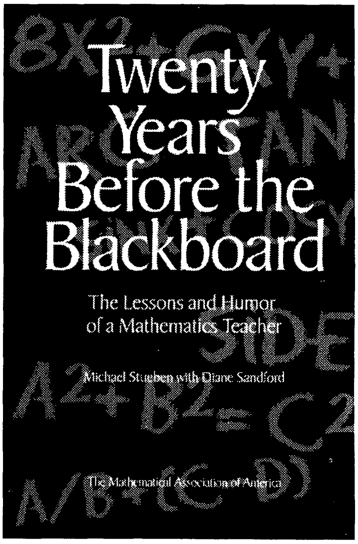
Contains the complete solutions and further discussion of nearly every exercise presented in the textbook. This includes both the mathematical/computational exercises as well as the research questions and investigations. Readers will benefit greatly from perusing solutions to the problems whether they have worked them out themselves or not. Students using this volume will still need to work out solutions of research questions using their own sources and adapting them to their own geographic locations, or using their own computational schemes, so this volume could well be useful for students in many course contexts. Enrichment material is included on the topics of some of the exercises. Advice for teachers who lack previous environmental experience, but who want to teach this material is also provided and makes it practical for such persons to offer a course based on these volumes.

Catalog Code: EVS/SA
150 pp., Paperbound, 1998, ISBN 0-88385-713-8
List: $18.95 MAA Member: $14.95

Phone in Your Order Now! ☎ *1-800-331-1622*

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789
or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

THE MATHEMATICAL ASSOCIATION OF AMERICA



Twenty Years Before the Blackboard

The Lessons and Humor of a Mathematics Teacher

Michael Stueben with Diane Sandford

Series: Spectrum

A perfect gift for the new teacher. . . or for anyone interested in the teaching of mathematics.

This book is the legacy of twenty years of mathematics teaching. During this time, the author searched for motivation techniques, mnemonics, insightful proofs, and serious applications of humor to aid his teaching. The result is this book: part philosophy, part humor, and part biography. Readers will be amused and enlightened on every page.

Mr. Stueben shows how he has used humor and wordplay to motivate his students. The book is filled with wonderful problems and proofs, as well as the author's insights about how to approach teaching problem solving to high school students. Sections of the book also treat the use of calculators and computers in the classroom. A section on mnemonics shows how teachers can use memory aids to help their students learn and retain material.

All in all, *Twenty Years Before the Blackboard* provides a goldmine of ideas for the classroom teacher. Although Mr. Stueben taught at the high school level, his book is an excellent "methods" book for mathematics teachers at all levels.

Read what Martin Gardner has to say about this fascinating book:

It's been decades since I read so entertaining a book about mathematics. The book is a treasure-trove of mathematical jokes, rhymes, anecdotes, word play, mnemonics, and beautiful proofs. For teachers there is an abundance of wise advice based on the author's twenty years in high school teaching. Mathematicians at all levels, from amateurs to college professors will not only chuckle over its gems, but learn much they did not know before. —Martin Gardner

Catalog Code: TYB/JR98
174 pp., Paperbound, 1998, ISBN 0-88385-525-9
List: $29.50 MAA Member: $23.50

Phone in Your Order Now! ☎ *1-800-331-1622*

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

Shipping and Handling: Postage and handling are charged as follows: **USA orders (shipped via UPS):** $2.95 for the first book, and $1.00 for each additional book. **Canadian orders:** $4.50 for the first book and $1.50 for each additional book. Canadian orders will be shipped within 10 days of receipt of order via the fastest available route. We do not ship via UPS into Canada unless the customer specially requests this service. Canadian customers who request UPS shipment will be billed an additional 7% of their total order. **Overseas orders:** $3.50 per item ordered for books sent surface mail. Airmail service is available at a rate of $7.00 per book. Foreign orders must be paid in US dollars through a US bank or through a New York clearinghouse. Credit Card orders are accepted for all customers.

Name ______________________

Address ______________________

City ____________ State _____ Zip ________

Phone ______________________

QTY.	CATALOG CODE	PRICE	AMOUNT
_______	TYB/JR98	_______	_______

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

Shipping & handling _______

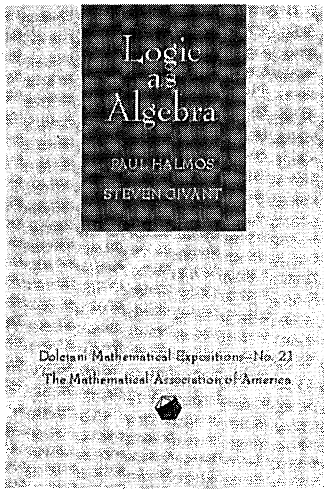
TOTAL _______

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ______________________ Expires ___/___

Signature ______________________

THE MATHEMATICAL ASSOCIATION OF AMERICA



Logic as Algebra

Paul Halmos and Steven Givant

Series: Dolciani Mathematical Expositions

This book is based on the notes of a course in logic given by Paul Halmos. This book retains the spirit and purpose of those notes, which was to show that logic can (and perhaps should) be viewed from an algebraic perspective. When so viewed, many of its principal notions are seen to be old friends, familiar algebraic notions that were "disguised" in logical clothing. Moreover, the connection between the principal theorems of the subject and well-known theorems in algebra becomes clearer. Even the proofs often gain in simplicity.

Propositional logic and monadic predicate calculus—predicate logic with a single quantifier— are the principal topics treated. The connections between logic and algebra are carefully explained. The key notions and the fundamental theorems are elucidated from both a logical and algebraic perspective. The final section gives a unique and illuminating algebraic treatment of the theory of syllogisms—perhaps the oldest branch of logic, and a subject that is neglected in most modern logic texts.

The presentation is aimed at a broad audience—mathematics amateurs, students, teachers, philosophers, linguists, computer scientists, engineers, and professional mathematicians. Whether the reader's goal is a quick glimpse of modern logic or a more serious study of the subject, the book's fresh approach will bring novel and illuminating insights to beginners and professionals alike. All that is required of the reader is an acquaintance with some of the basic notions encountered in a first course in modern algebra. In particular, no prior knowledge of logic is assumed. The book could serve equally well as a fireside companion and as a course text.

Contents: **What is Logic?:** To count or to think; A small alphabet; A small grammar; A small logic; What is truth?; Motivation of the small language; All mathematics. **Propositional Calculus:** Propositional symbols; Propositional abbreviations; Polish notation; Language as an algebra; Concatenation; Theorem schemata; Formal proofs; Entailment; Logical equivalence; Conjunction; Algebraic identities. **Boolean Algebra:** Equivalence classes; Interpretations; Consistency and Boolean algebra; Duality and commutativity; Properties of Boolean algebras; Subtraction; Examples of Boolean algebras. **Boolean Universal Algebra:** Subalgebras; Homomorphisms; Examples of homomorphisms; Free algebras; Kernels and ideals; Maximal ideals; Homomorphism theorem; Consequences; The representation theorem. **Logic via Algebra:** Pre-Boolean algebras; Substitution rule; Boolean logics; Algebra of the propositional calculus; Algebra of proof and consequence. **Lattices and Infinite Operations:** Lattices; Non-distributive lattices; Infinite operations. **Monadic Predicate Calculus:** Propositional functions; Finite functions; Functional monadic algebras; Functional quantifiers; Properties of quantifiers; Monadic algebras; Free monadic algebras; Modal logics; Monadic logics; Syllogisms.

Catalog Code: DOL-21/JR98
152 pp., Paperbound, 1998, ISBN 0-88385-327-2
List: $27.00 MAA Member: $21.95

Phone in Your Order Now! ☎ 1-800-331-1622

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

Shipping and Handling: Postage and handling are charged as follows: **USA orders (shipped via UPS):** $2.95 for the first book, and $1.00 for each additional book. **Canadian orders:** $4.50 for the first book and $1.50 for each additional book. Canadian orders will be shipped within 10 days of receipt of order via the fastest available route. We do not ship via UPS into Canada unless the customer specially requests this service. Canadian customers who request UPS shipment will be billed an additional 7% of their total order. **Overseas orders:** $3.50 per item ordered for books sent surface mail. Airmail service is available at a rate of $7.00 per book. Foreign orders must be paid in US dollars through a US bank or through a New York clearinghouse. Credit Card orders are accepted for all customers.

Name ______________________

Address ______________________

City ____________ State ____ Zip ______

Phone ______________________

QTY.	CATALOG CODE	PRICE	AMOUNT
______	DOL-21/JR98	______	______

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

Shipping & handling ______

TOTAL ______

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ______________ Expires ___/___

Signature ______________________

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036